PREDICTING SENSORY ATTRIBUTES FOUND IN A MODEL WINE USING SINGULAR VALUE DECOMPOSITION AND SUPPORT VECTOR MACHINES

By

DANIEL ARCHER DYCUS

A thesis submitted in partial fulfillment of
the requirements for the degree of

MASTER OF SCIENCE IN FOOD SCIENCE

WASHINGTON STATE UNIVERSITY
School of Food Science

JULY 2016

To the Faculty of Washington State University:

The members of the Committee appointed to examine the thesis of DANIEL ARCHER DYCUS find it satisfactory and recommend that it be accepted.

____

Carolyn F. Ross, Ph.D., Chair

____

Kevin R. Vixie, Ph.D.

____

Barbara Rasco, Ph.D.

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my committee who gave me this rare opportunity to follow my wildest curiosities concerning the predictive capacity of the sensory spectrum of wine. I especially would like to thank Kevin R. Vixie and his team who encouraged me through difficult times and the late nights playing games and programming with mathematicians Yunfeng Hu, and Joshua Kaminsky. My deepest gratitude is also due to Dr. Carolyn F. Ross for her acceptance and patience with outlying learners. I also want to thank Dr. Barbara Rasco who never penalized me for thinking outside the box and Dr. James Drake for molding my earliest theories, letting me crash his beloved Euclid, and our dialog about hyperplanes.

I want to thank my friends and family who helped me through hard times. My father and mother who patiently reminded me to persevere and helped me with the hurdles I thought were too high for me to ever overcome. Thanks to Charles Diako and my other Pullman family at The Black Cypress including the gentlemen Nick Foros, Nick Beymer, and Cory Preston who gave me a bicycle, food, and a place to live after hitch-hiking my way into the depths of the Pacific Northwest. I want to express my gratitude to my new family at Quixote in Napa. Once again, with no place to go, I was invited into the home of Aaron Pott for dinner. Eric Hagyard reminds me that great innovations are also often a matter of simple necessity.

Lastly, I would like to thank my fiancée Lidia Veronica Ruiz-Melara who reminded me to look up in my darkest hours so I could see the stars. She has been the greatest source of my inspiration and continues to encourage me with the patience of true love. Her encouragement has been the single largest contributor to the completion of this thesis. Te amo por siempre!

DEDICATION

To the Future…

PREDICTING SENSORY ATTRIBUTES FOUND IN A MODEL WINE USING SINGULAR VALUE DECOMPOSITION AND SUPPORT VECTOR MACHINES

Abstract

by Daniel Archer Dycus, M.S.
Washington State University
July 2016

Chair: Carolyn F. Ross

Advanced mathematical modeling can be applied to give insight into dynamic systems, including wine. Two of these mathematical models include Singular Value Decomposition and Support Vector Machines. This thesis used both SVD and SVR to data mine, examine overall sensory panel performance, and identify outliers from a previously conducted study on model red wines. In this previous study, twelve trained panelists rated the intensity of 20 different sensory attributes in model red wines varying in ethanol, tannin and fructose concentrations. These evaluation scores were analyzed using standard statistical methods including analysis of variance (ANOVA) and Principal Component Analysis (PCA). Our study went beyond these traditional methods and applied advanced mathematical modeling to these intensity attribute ratings thus predicting panelist perception based on the composition of the sample. Therefore, taking the data from the previous study and the known composition of the model wine, the thesis applied additional statistical and mathematical methods to create predictive models describing sensory perceptions.   Leaving out the panelists while re-running the model over multiple iterations with every combination of panelists removed, trained panelist who were considered “outliers” were identified using a “leave out then re-run” scenario. Using SVD and SVR, the sensory response of the wines could be predicted from the chemical composition of the wine, with prediction rates of prediction: r = 0.9077 (first order SVD), 0.9170 (second order), and 0.9245 (third order), respectively. The first order SVD corresponds to the optimal linear model. The SVR method did not perform as well (r = 0.9198) as the higher order approximations using SVD although, this was likely due to incomplete model optimization. Using the methods developed by this work, several future applications for datamining were contemplated, including smartphone app development for consumer preference, models to guide winemaker blending decisions, robotics and sensor array development, and identification of outliers who may represent a niche market.

TABLE OF CONTENTS

LIST OF TABLES

Page

LIST OF FIGURES

CHAPTER I:

INTRODUCTION

The ability to predict scientific relationships drives automation. In the current study, a statistical analysis was performed and a statistical learning theory method was applied to predict human intensity ratings from a trained sensory panel of a model red wine. This model red wine had already undergone sensory evaluation by a trained panel (Villamor et al. 2013). Using this sensory evaluation data, Singular Value Decomposition (SVD) was utilized to identify the fitness of our residual values and panelists who are considered to be "outliers." Higher order approximation was also performed on the SVD analysis to illustrate greater model fitness by increasing the number of terms thereby refining the approximation and increasing the model precision using a non-linear approach. A machine learning method used support vector machines regression (SVR) to create a linear regression model with parameter optimization. Both methods were then used to determine the feasibility of the respective model in predicting the sensory properties of the model wine based on its composition, including concentrations of ethanol, tannin, and fructose.

In a previous study (Villamor et al. 2013), thirty six (n=36) different model wine solutions were prepared based on a full-factorial design used to assess the effects of ethanol (0, 8, 10, 12, 14, and 16%, v/v), grape tannin, Biotan (500, 1000 and 1500 mg/L), and fructose (200mg/L and 2000mg/L). Eight chemical compounds were selected based on their sensory character and spiked at fixed concentrations: 250 mg/L 3-methyl-1-butanol (caramel), 0.002 mg/L dimethyl disulfide (sulfur), 1 mg/L 1-hexanol (herbaceous), 0.0001 mg/L 1-octen-3-one (earthy), 0.02 mg/L methoxyphenol (woody), 30 mg/L 2-phenylethanol (floral), 0.5 mg/L eugenol (spicy), and 0.03 mg/L β-damascenone (fruity). Twelve panelists participated in the trained sensory panel in which 10, 1 hour training sessions were conducted to ensure panelists evaluated the wines in a reproducible manner. The panelists rated the intensity of the above eight aromas, the corresponding eight flavors, as well as four mouthfeel descriptors including sourness, bitterness, heat, and drying. The sensory study was performed in triplicate. Our study only had access to data from the first and second repetitions of this experiment.

This study first used Singular Value Decomposition (SVD). SVD is a factorization of a real or complex unitary matrix. It is closely related to Principal Component Analysis (PCA) in that the eigenvectors of the covariance matrix in PCA are the same as the singular values computed in SVD. Numerous wine sensory studies have been performed in the past using PCA (Kwan et al. 1980, Schaefer et al. 1983, Guignard et al. 1987, Heymann et al. 1987, Noble et al. 1987, Heymann et al. 1989, Villamor 2012). However, thus far, few studies have explicitly utilized SVD. One study used SVD coupled with Artificial Neural Networks using Matlab for Windows version 4.2b (Sun et al. 1997). Our current study performed SVD on the matrix containing varying concentrations of ethanol, tannin, and fructose. From this analysis, outlying panelists were identified, at which point, we performed higher orders of approximation on the entire group of panelists using Matlab 2015a (Matlab 2015). The application of this method allowed a feasibility examination resulting in the prediction of intensity of sensory attributes based on the composition of the wine using a matrix factorization process. We also repeated the process with higher order approximations, and determined our outlying panelists by computing and recomputing the model fitness while removing panelists in a “leave one out” scenario.

Second, the study utilized a machine learning algorithm to see if a machine could learn to predict better than the SVD model. A Support Vector Machines regression (SVR) was selected using LIBSVM, a library for Support Vector Machines (Chang 2011). Support Vector Machines have been used in the past, achieving promising results when used to analyze a large sensory data set gathered from white and red wine sample evaluations from Portugal (Cortez et al. 2009). In the current study, Support Vector Machines was applied to construct a linear regression model (Support Vector Machines-Regression or just SVR) to predict if concentrations of ethanol, tannin, and fructose were responsible for intensity ratings of twenty attributes commonly associated with wine aroma, flavor, and mouthfeel. Optimization of the parameters was performed using the radial basis function kernel. This was written within a series of loops to determine minimal error, and provided as an optimized model output. The machine learning algorithm provided relative predictability of how a model wine tastes, smells, or feels in the mouth.

The overall research objective was to examine the predictability of perceived sensory responses based on the concentrations of ethanol, tannin, and fructose. Several methods commonly found in statistics and statistical learning theory were used to illustrate the prediction of the panelist response based on the composition of the model wine. This thesis is divided into 5 chapters. Following this introductory chapter is a review of literature on mathematical models used in sensory analysis in the past and the types of sensory tests used to gather this data. The third and fourth chapters contain the experiments where we performed SVD and SVR. Finally, a summary of conclusions and recommendations is included in chapter five.

CHAPTER II

LITERATURE REVIEW

Introduction to Wine Sensory Science and Datamining

The wine matrix, composed of thousands of compounds related to the biochemistry of the grape is vast and complex. Early attempts at classification of the compounds responsible for a sensorial response were first studied around 1959 in the United States of America at The University of California and from grapes coming from the Oakville research station. Here, a group of applied scientists and mathematicians attempted to make sense of sensorial reporting in wine. The vast body of work done by many academic researchers spans over four decades and corresponds with some of the oldest literature in the United States on the subject. This occurred as the wine industry in California, particularly Napa, began to flourish and thrive.

The new frontier and second largest producer of wine inside the United States is Washington State with its expansive Columbia River. In Washington, sensory evaluation caught on as well. Scientists there began the evaluations of wine grapes using methods which evolved over time and all across the globe. Today, comprehensive understanding of the sensory response system utilizes the knowledge of sensory testing methods from all over the world, the complex molecular system of wine components, as well as new advances in applied mathematics and data mining responsible for more than just wine research. The creation of customized analytical tools, tailored to the system itself is becoming one way to understand plausible theories about any given data set thus driving a black box approach for understanding the applications of new mathematical methods and techniques.

Sensory Testing and Psychophysics

In sensory evaluation, sensory scientists attempt to measure peoples' responses to foods or other consumer products. Psychophysics is the study of the relationship between energy in the environment and the response of the senses to that energy (Lawless 2013). In psychophysics, originally coined by Gustav Theodor Fechner, methods are classified into categories dealing with absolute thresholds, difference thresholds, scaling, and tradeoff relationships (Lawless, 2014). Several methods have been designed by sensory scientists in order to better understand and learn about human perception. Some of the methods used are simple tests presented to panelists which participate in environments like booths to either eliminate noise or to provide alternative noise levels like lighting (Stone et al. 2012). The panelist environment is an important source of variance (Stone et al. 2012). Sensory scientists carefully design testing facilities to control sources of background noise resulting in panelist variance. These tests, performed by sensory panelists include difference tests, quantitative tests, and hedonic tests. Examination of the results requires a careful regard to the use of scale, employing data analysis techniques such as perceptual mapping, multivariate tools, graph theory, and datamining (Lawless 2013). Mathematical methods for modeling are a crucial piece in understanding and extrapolating results. As scientists, a constant fight against variation requires a deep understanding of what variation is and why it occurs. To begin, sensory scientists start with choosing a sensory evaluation test.

Difference Testing

One of the many tools used in sensory testing is the difference test. These tests were initially known as Thurstonian models and have become increasingly complex over nearly one hundred years. Each augmentation of the tests or alteration in its design sought to eliminate assumptions made in the past (Bi 2007). Difference testing is used to determine if there is a difference among samples. Several types of major difference tests are commonly used in sensory testing. Some of the most common are paired comparison, duo-trio, triangle, tetrad, and r-index. Paired comparison determines if there is a difference between two samples. However, this test is not commonly used because it suffers from a response bias (Lawless & Heymann, 1998; M. O'Mahony, 1995). Duo-Trio tests consist of sets of three samples where one sample is identified the control (Fugelsang & Zoeklein, 2003; Larcher et al. 2008). Triangle tests also consist of sets of three samples; however, the control isn’t specified and the question asked of the panelists is “Which of these samples is different?”(Stone et al. 2012). Tetrad Testing provides more statistical power and requires four samples rather than three or two (O'Mahony 2013). This statistical power is better illustrated using an example of the Type II error. In many cases, one can argue that a major part of product research is more concerned with sameness than with differences. This creates an equivalence liability known as the error state of Type II error. In a nutshell, Type II error is that there really is a perceivable difference and that the sensory test missed the fact. Missing a difference that is really there; that is, failing to reject the null hypothesis when it is false. While most statistical research is aimed at preventing type I error or false positives, errors of type II are often much worse in applied product research and consumer product research. Type II represents missed opportunities (Lawless, 2013). The last type of difference testing is the R-index or R-type test. In this test, samples are compared to a standard and rated in one of four categories. When performing this type of difference testing, these categories are “standard”, “perhaps standard”, “perhaps not standard”, and “not standard”. The results are expressed in terms of R-indices which represent values of correct discrimination (Kilcast, 2003; Lawless & Heymann, 1998). This test is not commonly used, but has been applied for the determination of threshold values for caffeine (Robinson, Klien, & Lee, 2005) off-flavors in beef (An, Shim, Lee, Hong, & Lee, 2009), and wine (M. O'Mahony & Goldstein, 1986). While difference tests are commonly used to determine differences among samples, quantitative descriptive tests are also used to describe these specific differences.

Quantitative Testing

Quantitative tests are used to define the differences found in difference tests and can employ different scales. These scales include categorical, line scales, or logarithmic scales. Common tests in this type of test are the Spectrum Descriptive Analysis Method ™, Flavor Profile method, and descriptive tests. Quantitative descriptive analysis QDA profiles several different attributes at once (Lawless & Heymann, 1998). QDA has been used to characterize the aroma composition of Grenache rosé wines (Ferreira, Ortin, Escudero, Lopez, & Cacho, 2002), wines at different temperatures spiked with 4-ethylphenol (Cliff & King, 2009), and to evaluate Italian wines with an electronic tongue (Legin, Rudniskaya, Lvova, Vlasov, Di Natale, & D'Amico, 2003). Alterations to this type of test include manipulating the data by ranking and using non-parametric methods (Etievant, Issanchou, Ducruet, & Flanzy, 1989; Ugarte, Agosin, Bordeu, & Villalobos, 2005). Although this method compensates for variation among judges, sensory information can be lost.

Hedonic Testing

The last type of sensory test we'll cover is hedonic testing. Hedonic testing evaluates the affective reactions to products. Using choice methods and scaling methods, a consumer is asked to choose which product they like better, or to rate a set of products from least liked or accepted to most liked or accepted. Hedonic testing includes preference testing and acceptability testing, classically using the nine-point hedonic scale. The following nine categories are common and include: 9 = Like extremely; 8 = Like very much; 7 = Like moderately; 6 = Like slightly; 5 = Neither like nor dislike; 4 = Dislike slightly; 3 = Dislike moderately; 2 = Dislike very much and 1 = Dislike extremely (Lawless & Heymann, 1998). These tests require the panelists to be users of the product or product category, or sometimes simply purchasers of the product or product category. Hedonic testing has been used to sensory analysis of wine to examine quality, varietal, and regional reputation in New Zealand (Schamel & Anderson, 2003), to estimate price functions of Burgundian wine (Combris, Lecocq, & Visser, 2010), and to examine polyphenol preferences in wine (Noble & Lesschaeve, 2005). Hedonic tests are widely used in wine sensory science and examine preference with respect to scale.

Scale

One of the most important tools sensory scientists have at their disposal is scale. Scale gives insight into the intensity of the panelist response. Scale gives measurement by assigning numbers as symbols of properties and can be manipulated in accordance with the rules of mathematics (Gescheider, 1997). Scaling is a fundamental process of matching between two continua (Lawless, 2013). Numbers can either be continuous, as in the case of magnitude estimation, or discrete, as in the case of integer category scales. Sensory scientists use scales to help define blurred limits of resolution when dealing with panelist response (Lawless et al. 1998). It’s important to note that functions which fit log, exponential data, and linear functions provide a theoretical space where variants thereof are used to capture predictive evidence and provide additional insight into how panelists use and judge their sensorial responses. These functions of scale can be analyzed using perceptual maps, multivariate tools and graph theory.

Conclusions on Sensory Analysis

Sensory testing uses a variety of tests designed to capture panelist responses to sensorial data. As always, sensory scientists battle variation with the design of booths, the design of tests, and with mathematical constructs to mine their data. While sensory scientists utilize principles of psychophysics, the task of the sensory scientist is to apply sound scientific principles to capture valid data which can later be analyzed. This is done using difference testing, quantitative descriptions, hedonic testing, and scaling of their data. Each type of test uses different constraints to give the sensory scientists more insight into the interaction they are examining. Effects are commonly observed which illustrate attractiveness to certain variables, or synergisms. The understanding of statistical type I and type II errors is important as these errors are typically alleviated in opposing ways. In the next section, the historical approaches analysis in wine sensory science will be explored. While this is usually the last step in sensory science, it doesn’t need to be. New modelling techniques have already proven to be both valuable and insightful. Perceptual mapping, multivariate tools, graph theory, and statistical learning theory are at the edges of sensory science much like the outliers in data sets.

Mathematical Modelling Strategies in Wine Sensory Analysis

Sensory analysis of wine in the United States began around 1959 when a group of scientists and mathematicians from The University of California used difference tests to assess differences among wines (Amerine et al. 1959). They begin with paired tests, duo-trio tests, and the triangle test. Their panel took wines from two regions: one from the Oakville Research Station in Napa Valley and the other in Davis, California. Vintages 1958 and 1959 were used with eight panelists half of whom possessed more than 15 years of tasting and were considered expert tasters. The other panelists consisted of experience levels significantly lower than the first four and one panelist had little or no experience. Tasters were given a score sheet and asked to rate attributes on a scale of one to twenty. Analysis of Variance was then performed using a fixed effects model (Scheffé 1952). A later version of the analysis published in Hilgardia includes a test for homogenous variance and was done to modernize the methods performed previously (Ough et al. 1961). Mean scores were also calculated using the data (Duncan 1955). At this time, it appeared the model was poorly constructed because of mixed effects and heterogeneous variance. The fundamental model for analyses of variance was based on the assumptions of fixed effects and independent observations of equal variance while the authors continued to revisit the model and attempted to normalize the data, they found inexplicable variance coming from the panelists. The study concluded with the authors suggestion that more complicated models for the analyses of variance were required, especially the mixed models, but were not considered in detail because nearly all the unexplainable interactions could be reduced to a level of insignificance by omitting one aberrant taster or outlier (Ough et al. 1961).

The middle and early sixties of wine sensory analysis is led by a group of scientists and mathematicians in California. In 1966, this team of scientists and mathematicians published a paper concerning the predictability and grouping of sensory scores. In this study, they attempted to relate analytical scores to the sensory scores of four expert tasters. The analytical determinations included total acidity, volatile acidity, pH, extract, reducing sugar, ethyl alcohol, tannin, and color. The sensory test included an additional variable as the average score for all four judges. Thus, there were thirteen determinations on each wine, eight analytic, and five sensory (Baker et al. 1966). The results of this study reported a low predictability between chemical and sensory measurements, with high variation among wine vintages. The sensorial response was increasingly complex when they examined the interrelations among chemical and sensory determinations. The researchers also found distinct peaks among the correlation coefficients which influenced wine groupings, and interrelationships such as low-alcohol high-acid and high-alcohol low-acid wines. As shown in this study, grouping variables and multivariate analysis of categorical variables provided more insight into the predictability of the complex system. However, sensory analysis testing in wine still had a long way to go.

Over the next decade more and more scientists looked to sensory analysis when trying to describe and qualify wine character. In 1972, an article titled Recent Advances in Enology is co-authored by Amerine. In this review, a perceptual shift to how sensory analysis was perceived is noted. No longer were the tests simple ANOVA procedures meant to reveal statistical significance. At this point, ranking procedures were employed to assist in the classification of wines at an international judging (Amerine et al. 1972). This critical review states, “Statistical analysis of sensory evaluations of wine has been an American monopoly until recently.” German methods for sensory evaluation were developed and mentioned in the aforementioned review. The seventies illustrated a large increase in sensory studies of wine character.

Washington State University also began publishing wine sensory research around this time. In 1974 they published, “A Summary of Experimental Testing of Grape Varieties for Wine in Washington”(Carter et al. 1974). They used the same methods from the article in Hilgardia in 1961. One year later, California scientist Vernon Singleton wrote about the importance of using twelve expert tasters and the verification of data acquired from South African wines. In this study, researchers used a ten point scale for variables labeled: Desirable Aroma, and Astringency. They also use a twenty point scale for a grouped “quality” variable. Comment sections were also provided for the panelists, thus giving the researchers valuable insight into the score the panelists chose. The researchers utilized ANOVA to analyze the grouped variables and attempted to minimize panelist variance by reducing outliers through the use of expert panelists (Singleton et al. 1975). They used the Duncan’s multiple range test thus performing multiple comparisons of the grouped variables (Duncan 1955). One of the qualities of this test, a modification of the t-test, is to protect against false negative errors at the expense of an increased risk for false positives. The scientists included in their discussion that omitting particular samples increased the correlation coefficient or his model fitness from less than .50 to .55 or 55%. This meant just over half of his points were in the bounds of his statistically relevant union and illustrated a lack of fitness and apparent deviation from a linear model even after removing the outliers.

Wine sensory scientists seeking to use parabolic models possessing great fitness have been questionable since the earliest days of ANOVA when Scheffé designed the original construct. A respectful regard of the assumptions made is paramount to providing valuable insight into the analysis of sensory data. Attempts to reduce variance has been done in several ways. This was done by increasing the homogeneity of the scores presented by the panelists, eliminating erroneous samples, training expert panelists, and modifying the bounds of the statistical test. Finding ways to find models which fit the data better while respecting the assumptions of the model became a top priority for the scientists. One of the most intriguing aspects of fitting these early models included a discussion of predictability where the authors state no single factor is likely an indicator of maturity, and therefore of quality (Amerine et al. 1972). Again the conclusions coincided with a 1966 discussion, “Wine quality cannot be predicted from linear multivariate function of routine wine analyses using wines without distinctive varietal characteristics.” (Baker et al. 1966). The conclusions again pointed to the selected variables either as unimportant or the interrelationships between variables were of greater importance to the prediction of wine quality.

Later in 1973, scientists divided correlation of flavor with non-sensory data into two categories known as causal and predictive (Noble 1975) (Dravnieks et al. 1973). The causal approach was an attempt to identify and correlate compounds actually responsible for the particular sensory response under investigation. This approach had proven to be successful in model or very simple mixtures (Dravnieks et al. 1973). Interaction phenomena such as enhancement, masking and synergism in complex mixtures made the causal approach difficult (Dravnieks et al. 1973). The primary statistical method for locating indicators was, at this point, a stepwise regression analysis (Gianturco et al. 1974), (Hoff et al. 1975), (Moll et al. 1974). Much progress was made in the seventies in a search for predictive character using stepwise regression.

Then, in 1978, researchers attempted to predict panel preference for Zinfandel wine from analytical data (Ough et al. 1978). The wines were approximately nine months old at the time of preference evaluation. The crop level treatments included various levels of thinning and picking with a ripeness differences. Panelists were all enology students in their early twenties. Only three of them were female. The panelists received 2 days of training. They were asked to evaluate the wines on five scorecards, one each of aroma intensity, aroma preference, overall preference, taste intensity and taste preference. These rating cards utilized a 128 mm long line scale. The experiment used a randomized block design and a reference wine was used to minimize day to day variation. Each panelist replicated the aroma and flavor ratings three times and the taste ratings twice. Stepwise regression was used to develop the desired prediction equations (Dixon 1974). Then, analysis of variance for mean ratings of individual lots were made. The study illustrated that an individual wine could be identified correctly by the panelists 70% of the time. The panelist responses indicated significant differences in wine quality due to the treatment imposed. The researchers concluded the use of analytical data to predict wine quality was a real possibility although much more work would be required.

Using PCA in Wine Sensory Science

This work took an interesting turn in the eighties. During this decade, sensory scientists sought to solve problems arising from the assumptions of ANOVA, co-linearity, and heterogeneous variation by using more advanced mathematical methods. The results represented a closer look at the data and how relationships appear geometrically through the use of multiple dimensions and their relationships with the help of a rotational matrix while maximizing portions of sample variance. This decade illustrated the incorporation of principal component analysis (PCA) into the applied wine sensory science. Two researchers authored a paper in 1989 where they analyzed multivariate data using principal component analysis and canonical variate analysis (CVA). According to the authors, these methods were the most appropriate for examining relationships among variables and cases in which the data had a high ratio of between sample to within sample variance (Heymann et al. 1989).

This was not the first time PCA was used to analyze wine chemical data. It had also been performed to examine wine chemical data (Schaefer et al. 1983) and wine sensory data (Guignard et al. 1987, Heymann et al. 1987, Noble et al. 1987). It also was used in 1980 to analyze the sensory and chemical data of 40 Pinot noir wines. Here, researchers used PCA to investigate the correlations between chemical and sensory measurements (Kwan et al. 1980). CVA also is applied to interpret wine sensory data (Noble et al. 1984). Major differences exist between these two multivariate statistical techniques. PCA uses linear combinations of derivations from the original variables to explain the maximum amount of variation in the data set and are orthogonal or residing in a square matrix (Heymann et al. 1989). These principal components summarized the data with as little loss of information as possible (Mardia et al. 1979). Canonical Variate Analysis used linear combinations of the original variables selected to maximize the ratio of the between sample to the within sample variance. Canonical variates are not necessarily orthogonal and the actual angle between the canonical variates can be calculated (Tatsuoka 1971).

In PCA, several tests are used to determine the importance of the principal components indicating which number should be retained. These tests include the scree plot (Cattell 1966), the Kaiser criterion (Kaiser 1960), and the interpretability of the axes (Gnanadesikan et al. 1969). These tests were later found to provide insight into the system but not providing rigorous statistical tests of the significance of the axes (Heymann et al. 1989). In CVA, the significance of the axes was illustrated with Bartlett’s test (Green 1978, Chatfield et al. 1980). Also, confidence intervals were used to illustrate statistical significance among individuals (Chatfield et al. 1980).

The study in 1989 by Heymann and Noble analyzed twenty-one California Cabernet Sauvignon wines from four viticultural areas and rated them using descriptive analysis by thirteen trained judges. Eleven flavor terms were used in a PCA vs. CVA analysis. Fifty-eight California Chardonnay wines from three vintages were evaluated by ten trained judges, providing seven terms for PCA and CVA analysis. In both studies, each wine was evaluated twice by each judge. The researchers used a combination of FORTRAN programming and SAS for data mining. They concluded specific advantages to using PCA and CVA exist. Advantages of PCA included: its packaged into statistics packages as a tool for examining the relationships of a large number of variables and observe overall patterns in the data. Possible disadvantages of PCA include the lack of a statistical test to determine the number of important dimensions, as well as lack of a method to determine whether significant differences exist among the positions of the treatments in the sample space. However, the authors note CVA has an asset; the ability to determine the number of significant dimensions and to calculate confidence intervals for the treatments.

Statistical Learning Theory and Datamining

An interesting new way to examine problems with prediction now relies on statistical learning theory and machine learning. These techniques illustrated great progress in the nineties alongside what some might consider the beginning of artificial intelligence. The development of machine learning took a forefront in analysis with Singular Value Decomposition (SVD), Support Vector Machines (SVM), Artificial Neural Networks (ANN), Swarm Optimization, and Genetic Pare-to. These new models should be considered by enologists, analytical chemists, and sensory scientists as simple yet powerful new tools to be kept in mind when approaching sensory analysis.

Artificial neural nets were compared to results obtained by analysis with conventional Bayes discrimination analysis and Fisher discrimination methods (Sun et al. 1997). In this wine study, Principal component analysis was performed based on Singular Value Decomposition. Cluster analysis based on the Ward’s method is presented and followed by Bayes discrimination analysis and Fisher discrimination analysis. A leave one out scenario was used to increase the chances of finding and eliminating an outlier. The researchers concluded that principal component analysis and cluster analysis may be used to explore the rough structure of wine data, but neither method can completely recover the entire information included in the data set. The artificial neural network illustrated a discrimination rate of 100% for both training and prediction with the jackknife leave one out procedure due to its adaptability. Bayes stepwise discrimination method gave discrimination rates of 99.4% and 98.8% for training and prediction, respectively. Fisher’s discrimination method provided discrimination rates of 95.3% and 91.8% for training and prediction, respectively. The researchers concluded the use of an artificial neural network requires the longest amount of time to compute among all the methods. Also, both Bayes stepwise discrimination analysis method and the Fisher discrimination method can also be applied with slightly smaller discrimination rates, but with a shorter time to compute as compared with the neural networks.

The turn of the century represented major advances in analytical chemistry, computer science, and engineering. Using electronic noses and tongues, flavors, aromas and taste compounds can now be detected by machine. The persistent upgrades and augmentations of gas chromatography and mass spectrometry also illustrated greater ease of use and better separation of homogeneous mixtures. Traditional methods were updated with current head-space analysis research methods for capturing nearly volatile compounds associated with taste and flavor corresponding with several decades of studies (Stephan et al. 2000). Electronic noses consisting of sensor arrays use metal oxide thermal sensors, metal oxide conductimetric sensors, or bilayer lipid membrane coated mass sensors (Barnett et al. 1993, Kohl 1996, Mitrovics et al. 1997). E-tongues possessing voltammetric sensors on optimized sensor arrays were used to evaluate the complete antioxidant profile of red wines (Cetó et al. 2014, Cetó et al. 2014). Most recently, these analytical machines which possess a capability to produce large amounts of data were being coupled with kernels designed for feature selection and feature extraction. These kernels were powered by machine learning predictive algorithms.

In 2001, a study illustrating the use of artificial neural networks in regards to wine sensory science was published. This study mentioned the need to develop methods which would harness the power of artificial neural networks (Ferrier et al. 2001). The use of statistical learning theory in sensory science also coincided with an increase in the use of multivariate statistics. Multivariate statistics provided more insight into complex mixtures than univariate statistical methods such as simple linear regression. As instrumental analysis became more and more sensitive, capable of identifying and quantifying hundreds of compounds, pre-processing methods and data reduction became a necessity for those looking to acquire a glimpse at the larger picture. Partial Least Squares Regression became popular as increasingly large sets became apparent to the sensory scientists. While data reduction became a necessity, techniques became more robust and meaningful when a large number of cases were used. In trained sensory analysis, the labor involved to analyze 5 or 10 wines was enormous. A review article mentioned, most sensory studies do not use many samples to model relationships between sensory and instrumental data (Noble et al. 2002).

Partial least squares regression models continued to be developed during this time of the aforementioned review thus illustrating the predictability of red wine aroma properties from aroma chemical composition. In an article published in 2003, partial least squares was used to predict aroma properties from ionic signatures, with resulting correlation coefficients of 0.81 or 81% accuracy of prediction for the models (Aznar et al. 2003). The aromatic sensory characteristics of 57 Spanish aged red wines were determined by 51 experts inside the wine industry. The frequency of descriptors was used as a measurement of intensity. Gas chromatography mass spectrometry and flame ionization detectors were used to correlate descriptor intensity and the analytical machine data. Finding ways to incorporate least squares regression into data sets was deemed an intelligent way of looking at correlated vector spaces.

During this part of the early century, wine sensory scientists continued to look for insight into predictive models. In 2004, additional analytical methods were developed using chemometric principles like ionic signatures to drive characterization. These signatures were then loaded into artificial neural networks for classification. In one case, the wines were recognized 100% of the time and correctly predicted 78% of the time (Penza et al. 2004). The scientists concluded the results were significant and provided further insight to wine sensory scientists seeking to develop ways of distinguishing region specific identities through the use of artificial neural networks.

Up to this point, the use of multivariate statistics or multifactorial analysis has gone hand in hand or was often compared to partial least squares analysis, linear regression models, and principal component analysis. In 2009, another study was performed to predict wine character.

This study underlies an important principle for wine sensory scientists. As datasets became larger and more complex, there became many possible strategies for study. This allowed for data mining approaches to predict preference based on easily available analytical tests. The study in 2009 acquired a large data set of white and red wine samples from Portugal and applied three regression techniques under a computationally efficient procedure which performed simultaneous variable and model selection (Cortez et al. 2009). The support vector machine achieved promising results, outperforming the multiple regression and artificial neural network methods. The researchers concluded that such a model is useful to support the oenologist wine tasting evaluations and improve wine production. Furthermore, similar techniques can help in targeting markets by modelling consumer tastes from niche markets.

Support vector machines assisted the sensory scientist in understanding the boundaries of the product space. Knowing if two or more panelists share a theoretical space in terms of preference or perception is important because it can illustrate covariance and give insight into possible sources of variance. In one study, scientists created confidence intervals among the panelists using multiple factor analysis. In order to do this, they incorporated parametric bootstrapping (Dehlholm et al. 2012). Bootstrapping uses random sampling with replacement. This allowed the scientist to view measures of accuracy. Similar to confidence bounds found in statistics in a variety of tests or metrics, the parametric bootstrapping method assigned confidence ellipses to the data already analyzed using multiple factor analysis. The researchers use R as their programming environment. They found this approach was suitable for generating an overview of product confidence intervals and also applicable for data obtained from ‘one repetition’ evaluations. Furthermore, they concluded bootstrapping is a convenient way to get an overview of variations in different studies performed on the same set of products and the graphical display of confidence ellipses eased interpretation and communication of results.

Later in 2012, newer datamining methods were incorporated where sensory scientists use genetic programming, statistical techniques, and swarm optimization. Genetic programming models were first eluded to for wine sensory research in 2001 in a discussion on artificial neural networks (Vlassides et al. 2001). Swarm optimization is known to trap into local minima similar to support vector machines. Swarm optimization models give valuable insight into vector fields and the subspaces of data sets because when an individual is close to the optimal particle, its velocity will approximate to zero (Leifu Gao 2009). In the study in 2012 researchers attempted to datamine sensory evaluation data through extreme sparsity and a large variation in responses from panelists (Veeramachaneni et al. 2012). Their approach employed genetic programming (symbolic regression) and ensemble methods in order to generate multiple diverse explanations of assessor liking preferences with confidence information. Using the produced ensembles to unobserved regions of the flavor space, statistical techniques were used to extrapolate and then segment the panelists into groups possessing a propensity to like flavors. Finally, a two-objective swarm optimization was applied to identify flavors which were well and consistently liked. The researchers defined a new space which respects the evidence that the response and explanatory variable relationship differs among panelists and exploits rather than inaccurately averages the differences. This method postponed decision making regarding a prediction and decision boundary until the end of the analysis, an approach that was not used by historical modelling approaches. Finally, as macro-level behavior emerged and more was known about the panelists, decision boundaries were rationally imposed on the probability space, allowing for segmentation. The researchers concluded an affirmation of genetic programming symbolic regression methods which has since evolved into a mature field. In this case, genetic programming allowed the researcher an ability to decompose high variation into a sequence of solvable problems.

The most recent research on predicting sensory character from wines includes the use of statistics, linear algebra, and machine learning. Multivariate methods recently developed illustrate classification of wines based on multiple variables to create signatures and how they can be used to infer properties found in wines from a particular region (Shanmuganathan et al. 2013, Selih et al. 2014). Linear algebra constructs give sensory scientists insight into discriminating classifiers of set spaces (Wang et al. 2014). Machine learning allows sensory scientists not only the ability to discriminate signatures but also to predict complete products based on multiple signatures coming from analytical machines (Cetó et al. 2014, Gómez-Meire et al. 2014, Hosu et al. 2014, Selih et al. 2014, Tao et al. 2014, Wang et al. 2014). The power of machine learning, analytical machines, and sensory sciences is beginning to take a strong foot hold in applied food sciences. Datamining and mathematics serve as principles for understanding problems in complex data sets and provide insight through the use of predictive algorithms. In this study, we attempt to incorporate machine learning, prediction, and linear algebra constructs to provide ways for wine sensory scientists to find truths about their datasets without relying on assumptions.

CHAPTER III

SINGULAR VALUE DECOMPOSITION

Introduction to Singular Value Decomposition and Our Data

The singular value decomposition (SVD) is a factorization of a matrix into a product of matrices (Equation 1). The decomposition of the original matrix ($\mathbf{M}$) is a product of an orthogonal matrix ($\mathbf{U}$), a non-negative diagonal matrix ($\mathbf{\Sigma}$), and the transpose of a second orthogonal matrix ($\mathbf{V}^*$). The SVD finds directions along which matrix multiplication is equivalent to scalar multiplication but has greater generality than eigenvector-eigenvalue decompositions since the original matrix need not be square.

$$\mathbf{M} = \mathbf{U\Sigma V}^* \qquad \textbf{Eq. 1}$$

In two dimensions, the decomposition is visualized in three parts (Figure 1). The first part is an initial rotation of the domain of $\mathbf{M}$. The second part is a direction dependent scaling of space using a diagonal matrix. The third part is a second rotation. The process is illustrated most simply in a 2×2 matrix.

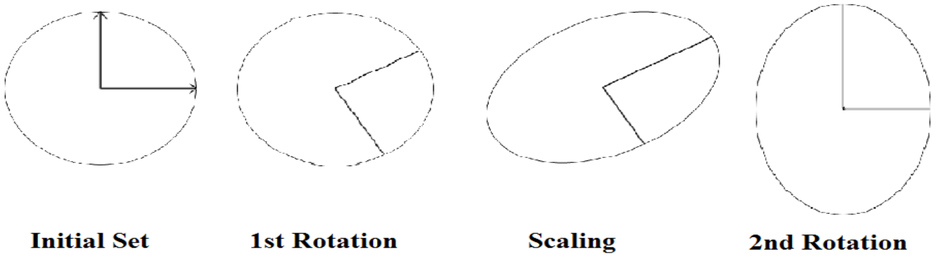

Figure 1: SVD

This experiment used Singular Value Decomposition (SVD) to predict how well each measured sensory attribute (n=20) can be predicted by concentrations of ethanol (n=6), tannin (n=3), or fructose (n=2). First, two matrices were imported from the previous study into Matlab version 2015a (Mathworks, Natick, MA). One matrix, named $\mathbf{A}$, contained attribute intensity ratings (864x20). A second matrix, named $\mathbf{C}$, contained concentrations of ethanol, tannin, fructose, and a column of ones used as a bias constant (864x4). These matrices had individual panelist data extracted from them. SVD was then computed for each of the twelve panelists. $\mathbf{C}_{\boldsymbol{i}}$, where $\boldsymbol{i}$ was used as an index value and refers to the panelist number or individual is simply the 72 rows of $\mathbf{C}$ that came from panelist $\boldsymbol{i}$. Similarly, $\mathbf{A}_{\boldsymbol{i}}$ is the 72×20 matrix that corresponds to the 72 rows in $\mathbf{A}$ that come from panelist $\boldsymbol{i}$.

Computing the SVD of $\mathbf{C}_{\boldsymbol{i}}$, $\mathbf{C}_{\boldsymbol{i}} = \mathbf{U}_{\boldsymbol{i}}\mathbf{\Sigma}_{\boldsymbol{i}}\mathbf{V}_{\boldsymbol{i}}^{*}$ where $\mathbf{U}_{\boldsymbol{i}}$, is a 72×72 real matrix where the 72 columns are the left singular vectors. $\mathbf{\Sigma}_{\boldsymbol{i}}$, is a 72×4 rectangular diagonal matrix with non-negative real numbers on the diagonal known as the singular values of $\mathbf{C}_{\boldsymbol{i}}$ and represented a scaling of the data. Finally $\mathbf{V}_{\boldsymbol{i}}{}^{*}$, is the transpose of $\mathbf{V}_{\boldsymbol{i}}{}^{*}$, a 4×4 orthogonal matrix whose columns are known as the right singular vectors.

Introduction to Subspaces and Projections

Now that the factorization is complete, a brief explanation of how it's utilized is illustrated with a simple example. First, take a 3-dimensional coordinate space with x, y, and z coordinates. Now let's place two vectors which both start at the origin. These vectors are named $b_1$and b_2 (Figure 2).

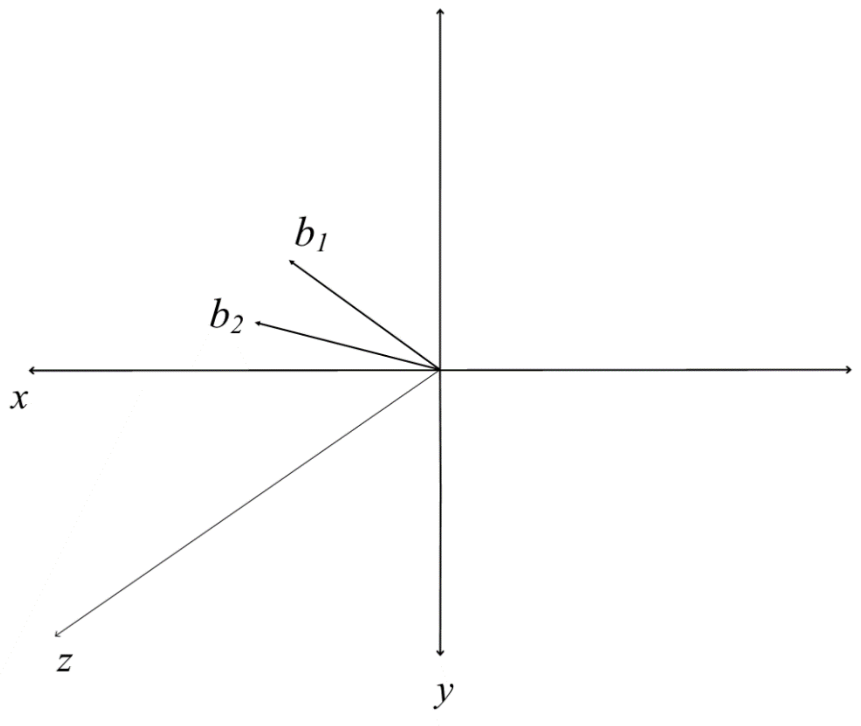

Figure 2: Two vectors that start at the origin.

The vectors described share a space which can be referred to as a linear subspace. The subspace, labelled *P* is the span of vectors b_1 and b_2 (Figure 3).

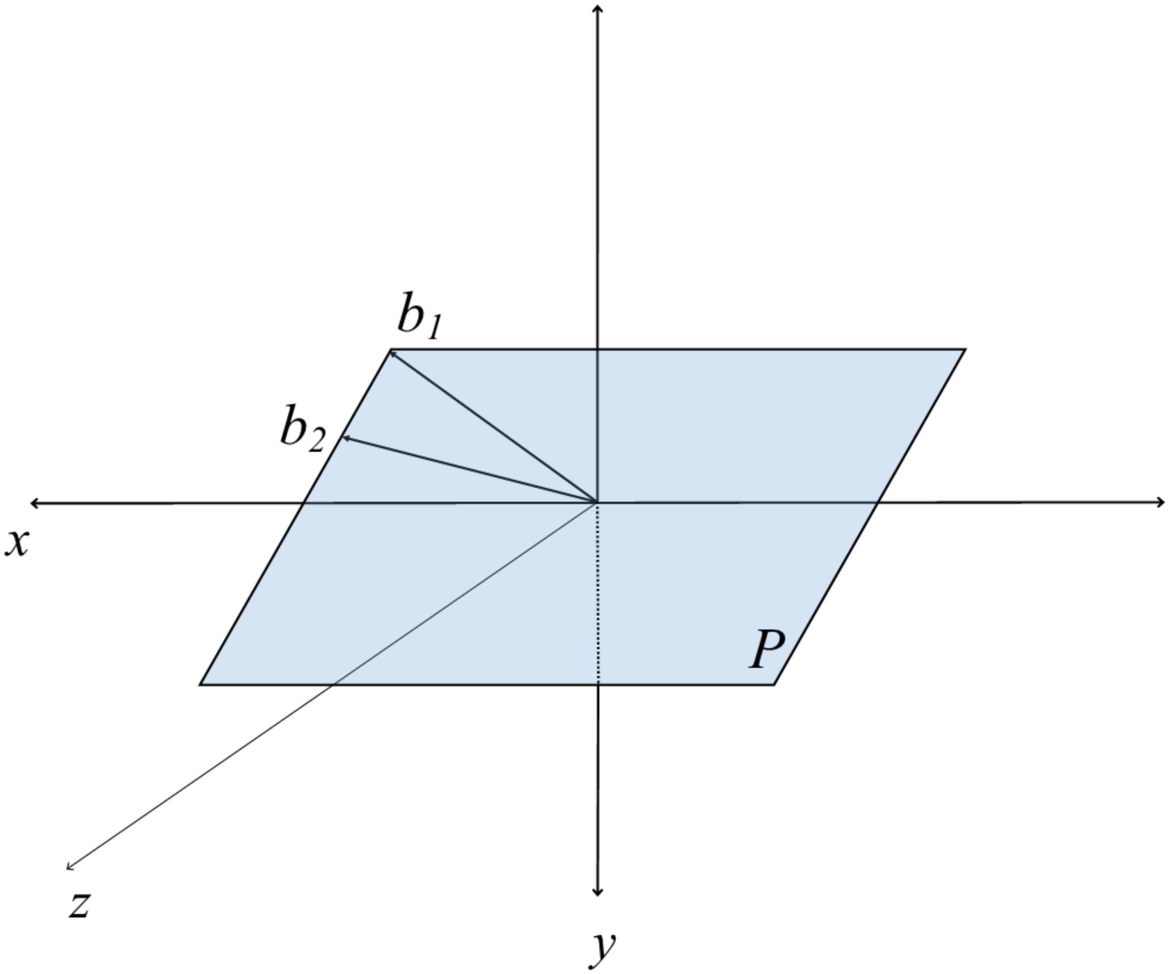

Figure 3: The span of two vectors

Once the subspace has been created a new concept of projection is introduced. The projection of a vector *v* onto *P* is the closest point *w* which lies on *P*, the subspace created by the space spanned by vectors b_1 and b_2 (Figure 4).

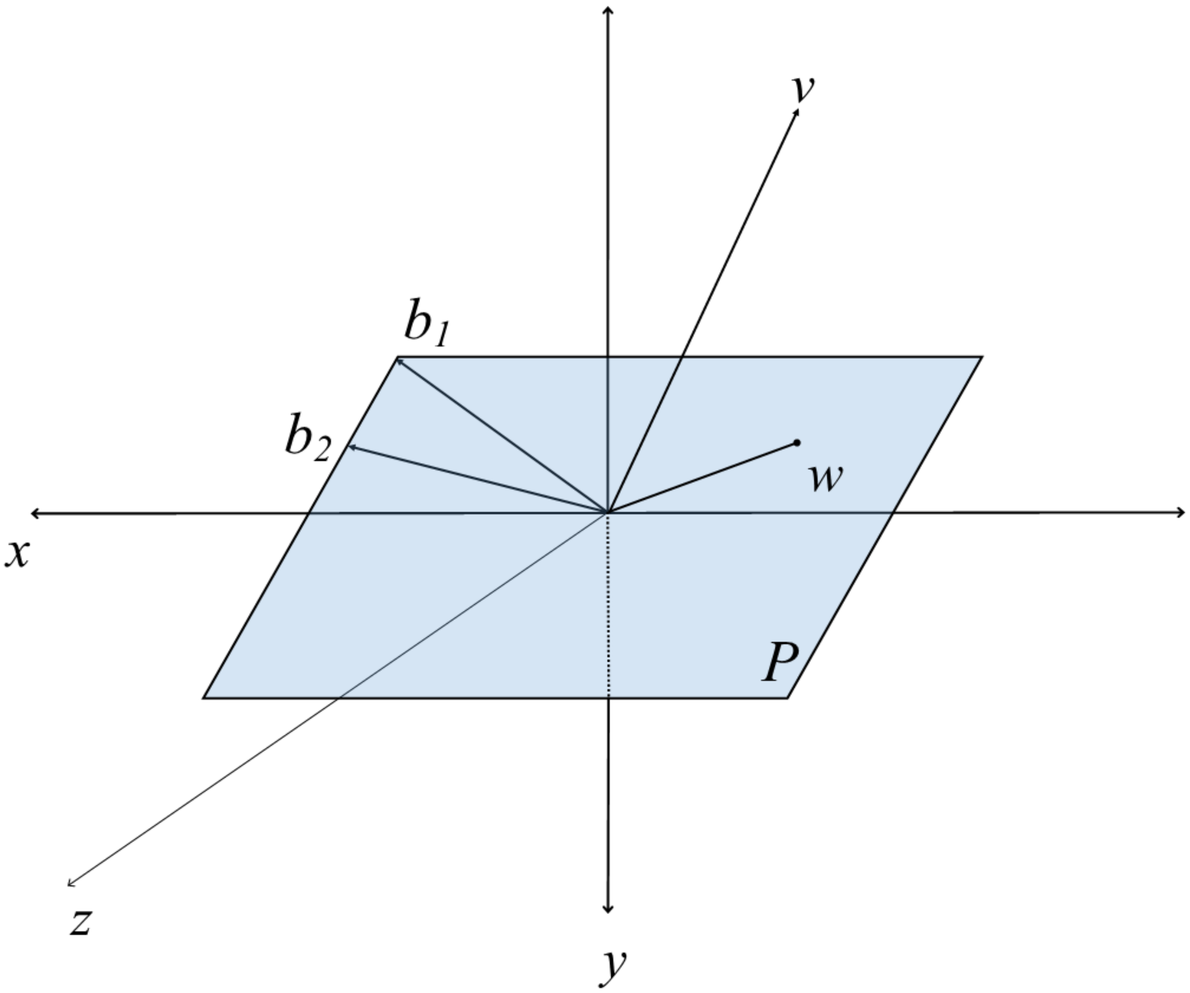

Figure 4: The projection

Finally, after creating the subspace and projecting v onto the plane spanned P an error vector is calculated. The error vector is $(v - w)$ and the error is the length of this error vector.

This error refers to the error resulting from an attempt to approximate v with w, which is the best approximation to v in the linear subspace P.

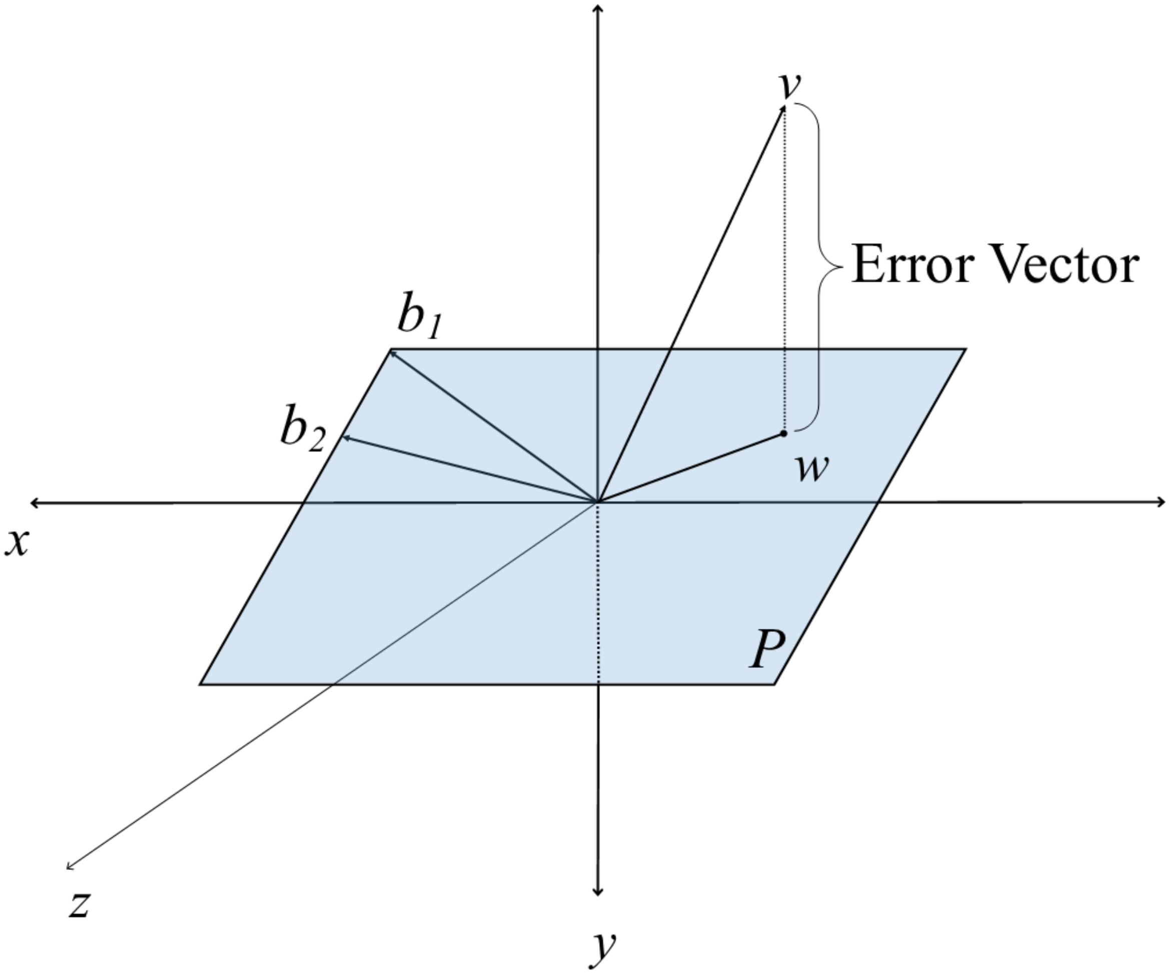

Figure 5: Error Vector

Methods of Analysis

In this experiment instead of being 2 dimensional, the plane P, is a 4 dimensional plane that is the span of the 4 columns of $\mathbf{C}_{\boldsymbol{i}}$**.** The space in which the plane resides is 72 dimensional, which is the dimension of each of the column vectors which make up $\mathbf{C}_{\boldsymbol{i}}$. So instead of a 2 dimensional plane in 3 dimensional space, there exists a 4-dimensional plane in 72-dimensional space ($\boldsymbol{C_i}$). Now, one of the columns of $\mathbf{A}_{\boldsymbol{i}}$**,** which is named $\boldsymbol{a}$ is projected onto the subspace P, spanned by the columns of $\mathbf{C}_{\boldsymbol{i}}$. The projection at $\boldsymbol{a}$ onto P is now easily computed using the SVD of $\mathbf{C}_{\boldsymbol{i}}$**.** This is done in two steps.

$$\boldsymbol{\beta} = \mathbf{V}_{\boldsymbol{i}}\boldsymbol{\Sigma}_{\boldsymbol{i}}{}^{\mathbf{-1}}\mathbf{U}_{\boldsymbol{i}}{}^{*} * \mathbf{A}_{\boldsymbol{i}} \qquad \textbf{Eq. 2}$$

Where $\boldsymbol{\Sigma}_{\boldsymbol{i}}{}^{\mathbf{-1}}$ is the pseudoinverse of $\boldsymbol{\Sigma}_{\boldsymbol{i}}$**.** Second, the projection $\boldsymbol{a_P}$ computed.

$$\boldsymbol{a_P} = \boldsymbol{C_i}\boldsymbol{\beta} \qquad \textbf{Eq. 3}$$

Note that β has four components: the coefficients of the columns of the columns of $\boldsymbol{C_i}$ that yield the projection. They are the parameters of the optimal linear model combining ethanol, tannin, fructose and an offset to get a prediction of the attribute corresponding to the column $\boldsymbol{a}$.

Analysis continues with the creation of an error vector. $\mathbf{C}_{\boldsymbol{i}}\boldsymbol{\beta}$ is the projection of $\boldsymbol{a}$ onto the span of the columns of $\mathbf{C}_{\boldsymbol{i}}$, and $\boldsymbol{\beta}$ therefore gave the best possible linear model for the attribute corresponding to $\boldsymbol{a}$. This forms an error vector where the equation is listed below.

$$\boldsymbol{e} = \mathbf{C}_{\boldsymbol{i}}\boldsymbol{\beta} - \boldsymbol{a} \qquad \textbf{Eq. 4}$$

This error, represents the difference of a perceived human sensory response $\boldsymbol{a}$ and the the optimal linear model using the concentrations of ethanol, tannin, fructose, and a vector of ones.

If the relationship between attributes and chemical composition is linear, one would expect no error. If there is noise in the data, the error would be commensurate with the size of this noise. Larger amounts of error may be the result of non-linear relationships between stimulus and response

The point of this exercise was to test the hypothesis that each individual possessed a linear function which converts the chemical composition found in wine to the intensity of a perceived sensory attribute. As a measure of linearity we could use the normalized error.

$$\boldsymbol{\varepsilon} = \frac{\|\boldsymbol{e}\|^2}{\|\boldsymbol{a}\|^2} \qquad \textbf{Eq. 5}$$

In this section because we are using projections, we can also use

$$\boldsymbol{r} = \mathbf{1} - \boldsymbol{\varepsilon} \qquad \textbf{Eq. 6}$$

Which corresponds to the square of the correlation coefficient. When $\varepsilon \to 0$ $(r \to 1)$ then we conclude the model is linear or very close to linear.

Results and Discussion

This analysis used Singular Value Decomposition and least squares to determine if a linear function can adequately describe the intensity rating of wine attribute from a series of molecular compositions. The analysis produced 240 linear predictions of $\boldsymbol{\beta}$. Then we compute the coefficient $\boldsymbol{r}$, which measured how well a linear model describes a sensory function.

The sensory data coming from the panel was grouped as matching aromas and flavors with a final category called mouthfeel attributes. Significant research has been done on the shared receptors in the mouth and nose (Bakalar 2012). This physiology was found to be highly positively correlated with one another in a preliminary study where SVD was performed using only the aromas and flavors from this dataset (Table 1). Furthermore, the sensory evaluation panelists performed better as a whole at producing consistent intensity ratings among 4 mouthfeel attributes over the aromas and flavors (Figure 6).

Upon examination of the panelists as individuals, it became apparent panelists three, six, and eleven contribute the most variation to the set as a whole (Figure 7). These panelists represent values which might be considered outliers from the group, with low amounts of linear fitness. While we kept these panelists in the data set, we want to briefly make note of them.

The outliers themselves bring about an interesting conversation about why they chose intensity ratings which were unlike the majority of the other panelists. This could be the result of sensory testing bias, anosmia, or a non-linear function (Lawless et al. 1998, Stone et al. 2012) between the wine attribute and the chemical composition of the wine. The major problem with a linear approach to modelling this system lies in the suspicion that wine is a non-linear and dynamic system. Much of what has been found in nature is non-linear and therefore other approaches for modelling should be considered. In order to find better solutions for predicting panelist responses to the chemical composition of wine, non-linear methods must also be developed.

Table 1: Trained panelist r- values from SVD Aromas and Flavors

Caramel	Earthy	Floral	Fruity	Vegetal	Spicy	Sulfur	Woody
.925	.937	.936	.940	.950	.920	.887	.933

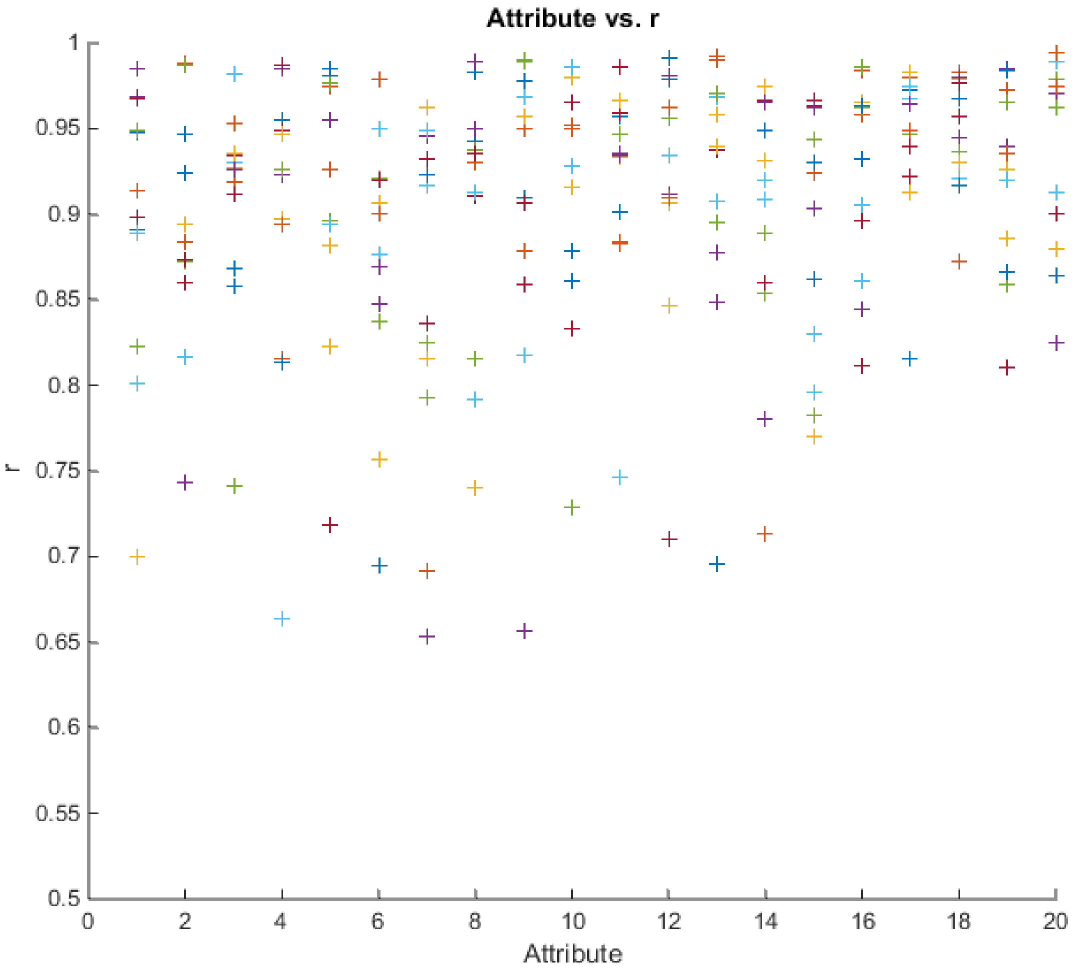

Figure 6: Attribute vs. r-value from trained panel

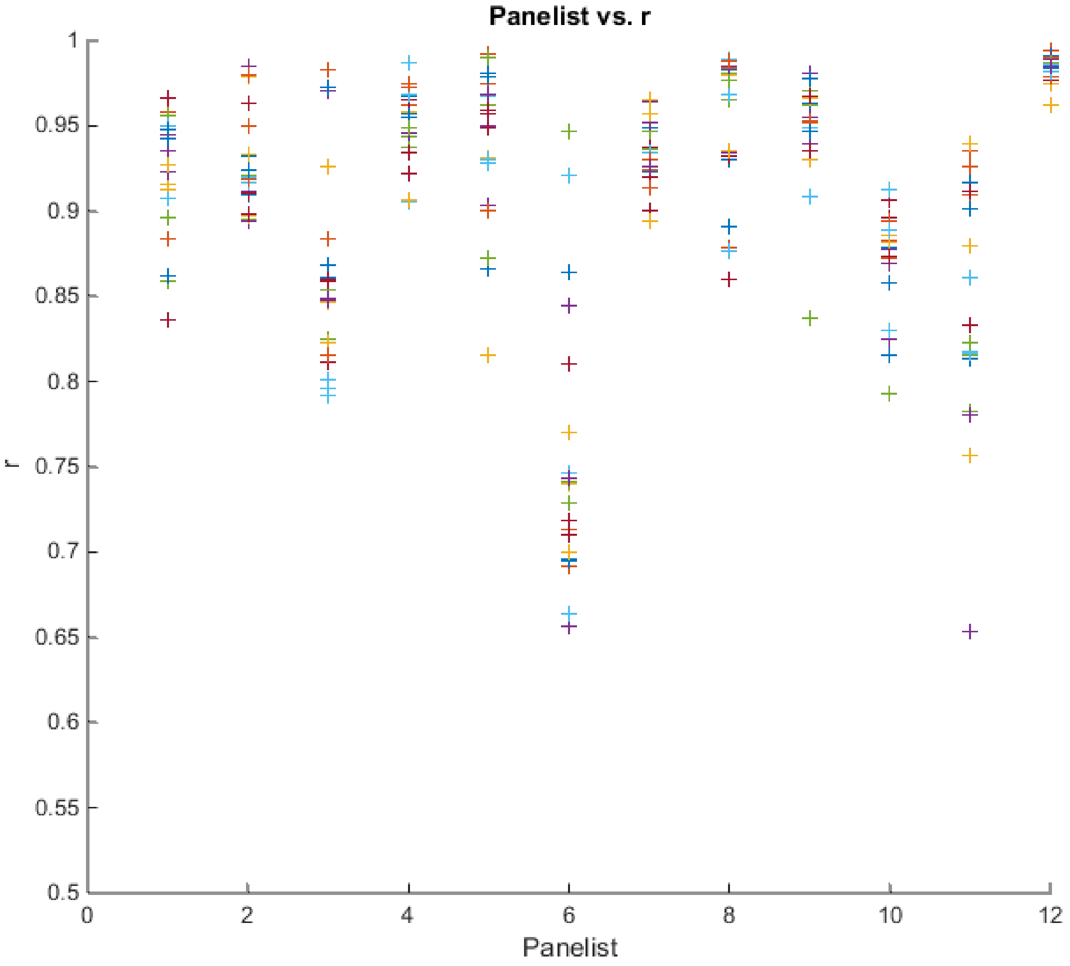

Figure 7: The range of r-values for each trained panelist

The discussion of our analysis warrants a secondary discussion of the fitness of our data and the inclusion of additional orders of approximation. The continuation of the SVD portion of our experiment includes second and third order terms. Second order approximation is defined as a quadratic polynomial with a degree of two and is represented geometrically by a parabola. Third order approximation is generally referred to as a polynomial interpolation and is represented geometrically with a figure which looks similar to that of a chair. Performing the higher orders of approximation on the data which is transformed using the rules of SVD results in greater linear fitness from our linear model. The following illustration represents how approximation works with a simple sine wave, the first approximation which is simply a line which captures the slope of the sine function, then our third and fifth order terms which capture more of the sine function (Figure 8).

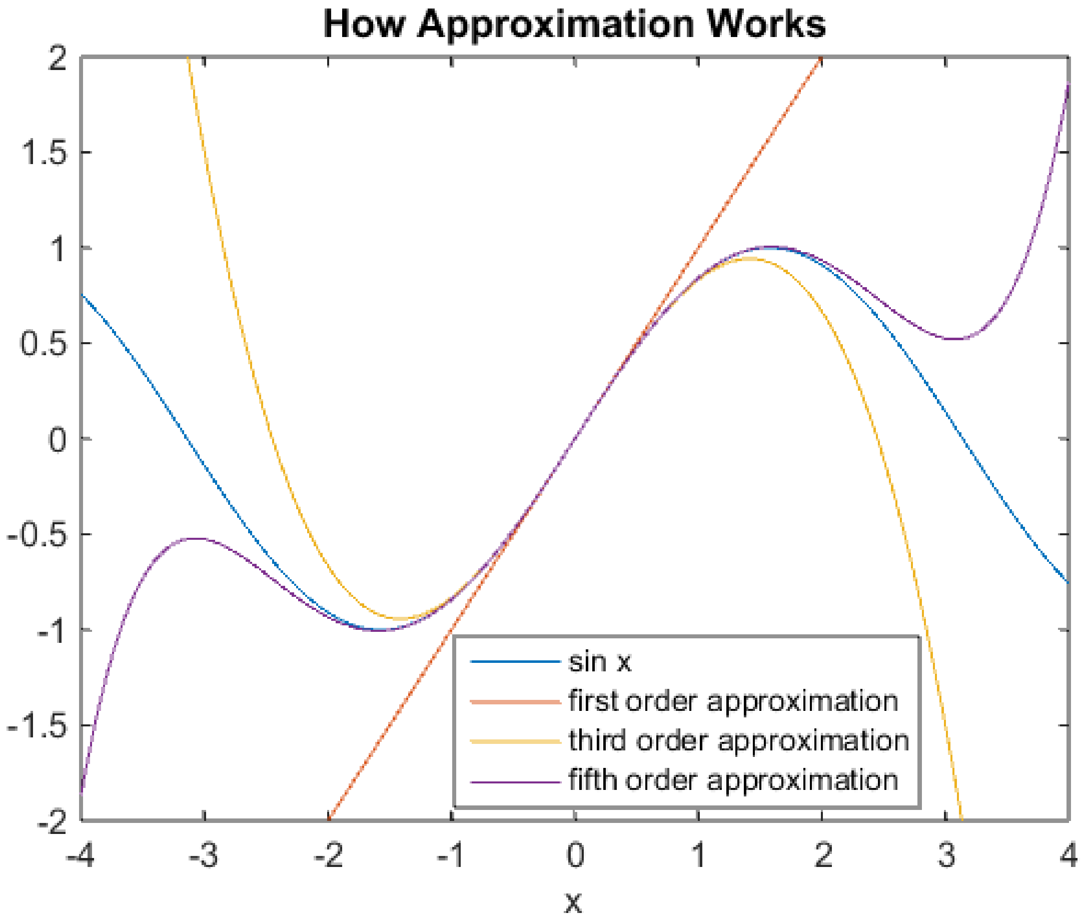

Figure 8: Approximation of a sine wave

The initial dataset, which was decomposed and transformed, contained a four column matrix which included concentrations of ethanol, tannin, fructose, and a column of ones (Equation 7). When performing the second order approximation, we created a new matrix which included the square of each of the concentrations, the second order factors, our initial concentrations, and the column of ones so we can preserve our bias constant. The third order approximations were done in the same manner by cubing our terms, maintaining our factors, and retaining our initial data and the column of ones. The equations below illustrate the construction of the matrices used in first, second, and third order approximation. Here, *C* is the concentration matrix for the SVD, *E* is the concentration of Ethanol, *T* is the concentration of Tannin, and *F* is the concentration of Fructose. The subscripts represent the 72 values reported by each panelist.

The first order approximation matrix:

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{E_1} & \boldsymbol{T_1} & \boldsymbol{F_1} & \boldsymbol{1_1} \\ \vdots & \vdots & \vdots & \vdots \\ \boldsymbol{E_{72}} & \boldsymbol{T_{72}} & \boldsymbol{F_{72}} & \boldsymbol{1_{72}} \end{bmatrix} \qquad \textbf{Eq. 7}$$

The second order approximation matrix **(Equation 8)**:

$$C = \begin{bmatrix} (E_1^2) & (T_1^2) & (F_1^2) & (E_1 \cdot T_1) & (E_1 \cdot F_1) & (T_1 \cdot F_1) & (E_1) & (T_1) & (F_1) & (1_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (E_{72}^2) & (T_{72}^2) & (F_{72}^2) & (E_{72} \cdot T_{72}) & (E_{72} \cdot F_{72}) & (T_{72} \cdot F_{72}) & (E_{72}) & (T_{72}) & (F_{72}) & (1_{72}) \end{bmatrix}$$

Note that instead of a subspace spanning four columns, now the subspace is spanned by ten.

The third order approximation matrix **(Equation 9)**:

$$C = \begin{bmatrix} (E_1^3) & (T_1^3) & (F_1^3) & (E_1^2 \cdot T_1) & (E_1^2 \cdot F_1) & (T_1^2 \cdot E_1) & (T_1^2 \cdot F_1) & (F_1^2 \cdot E_1) & (F_1^2 \cdot T_1) & (E_1 \cdot T_1 \cdot F_1) & (E_1^2) & (T_1^2) & (F_1^2) & (E_1 \cdot T_1) & (E_1 \cdot F_1) & (T_1 \cdot F_1) & (E_1) & (T_1) & (F_1) & (1_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ (E_{72}^3) & (T_{72}^3) & (F_{72}^3) & (E_{72}^2 \cdot T_{72}) & (E_{72}^2 \cdot F_{72}) & (T_{72}^2 \cdot E_{72}) & (T_{72}^2 \cdot F_{72}) & (F_{72}^2 \cdot E_{72}) & (F_{72}^2 \cdot T_{72}) & (E_{72} \cdot T_{72} \cdot F_{72}) & (E_{72}^2) & (T_{72}^2) & (F_{72}^2) & (E_{72} \cdot T_{72}) & (E_{72} \cdot F_{72}) & (T_{72} \cdot F_{72}) & (E_{72}) & (T_{72}) & (F_{72}) & (1_{72}) \end{bmatrix}$$

Note that instead of a subspace spanning ten columns, now the subspace is spanned by twenty.

Performing the higher orders of approximation on the data which is transformed using the rules of SVD results in greater linear fitness from our linear model. The mean r-squared improvement is nearly 2% in the third order approximation and around 1% in the second order approximation. The mean r-squared improvement is between 2% and 5% among the panelist outliers. The figure on the top of the next page illustrates the improvement of the panelist r values among each of the twenty attributes.

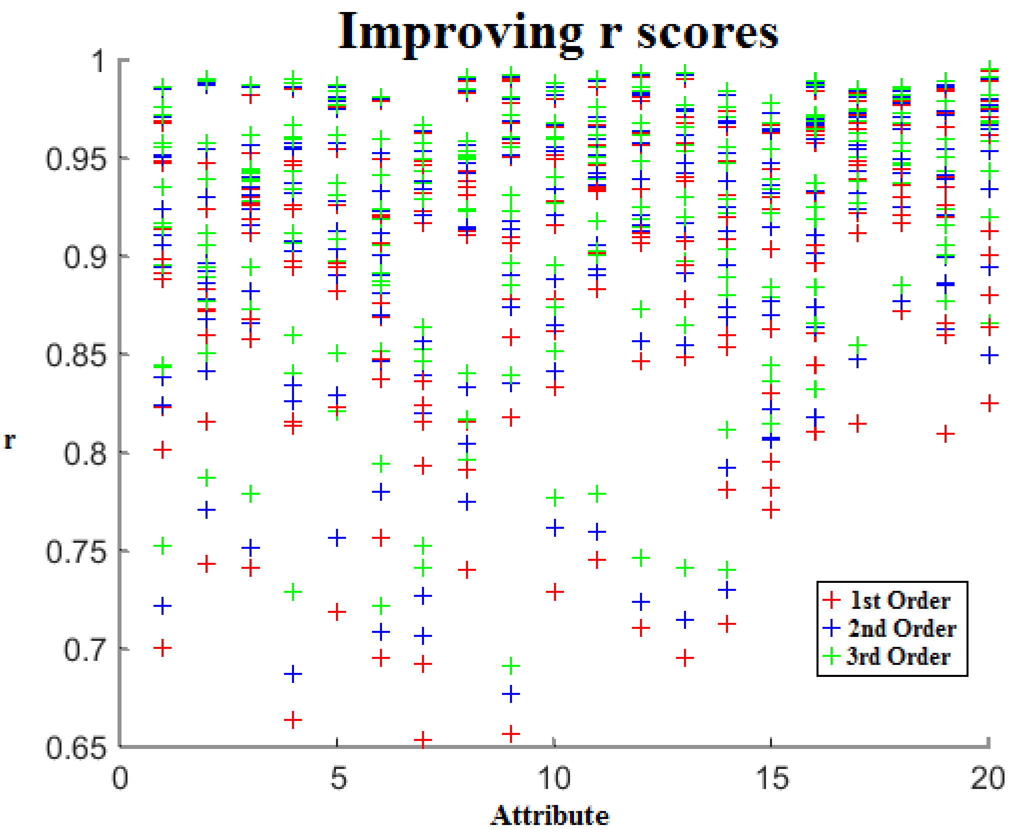

Figure 9: Improving range of r-values for each panelist

Conclusions

In the latter portion of our discussion, we followed the linear SVD portion with a more in-depth examination using higher orders of approximation. The once linear regression model is now a non-linear model. Using higher orders of approximation meant that we expected to see an increase in model fitness. Our hypothesis was correct in this case as we saw improvement of our r-squared values. The non-linear methods are able to increase fitness by capturing more of the data, thereby increasing the r-square value. This method is preferred as we gain model improvement while maintaining our panelist numbers. We continue to work with non-linear models in the next section where we use Support Vector Machines, an application of statistical learning theory.

CHAPTER IV

SUPPORT VECTOR MACHINES

Introduction to Support Vector Machines

Support Vector Machines (SVM) is a supervised learning model or machine learning method designed to classify data. To illustrate how SVM works, we first take a set of training examples, each marked as belonging to one of two categories. The SVM training algorithm builds a model that assigns new examples into one category or the other. In addition to performing linear classification, SVM can perform a non-linear classification using kernels which implicitly map the inputs into high-dimensional feature spaces. SVM relies on the use of boundaries and the creation of a decision rule. Here, we follow Winston’s exposition closely (Vapnik 1998, Winston 2010).

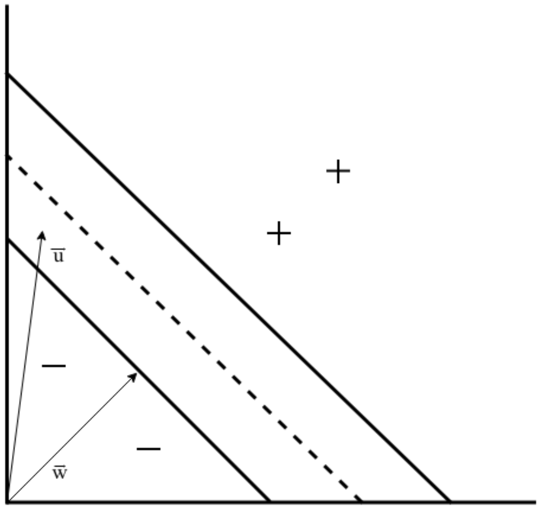

Figure 10: Creating a decision rule

The initial problem is determining the separable distance between the two classes either (+) or (-) in our case (Figure 10). This can be done by first creating a vector which is perpendicular to the street; $\vec{w}$. Using an unknown we want to know if a said point is on the right side or left side of our street (Figure 10). We can do this simply multiplying the two vectors in Equation 10.

$$\vec{w} \cdot \vec{u} \geq c \qquad \textbf{Eq. 10}$$

$$\text{or } \vec{w} \cdot \vec{u} + b \geq 0 \; when\, positive$$

$$where\; c = \; -b$$

Performing this simple multiplication has given us a decision rule where we illustrate the location of our street but we also need constraints for this rule. We define our constraints for the decision rule in Equation 11.

$$\begin{aligned} &\vec{w} \cdot \vec{x} + b \geq 1 \\ &\vec{w} \cdot \vec{x} + b \leq -1 \end{aligned} \qquad \textbf{Eq. 11}$$

The addition of another variable makes things more convenient.

$$y_i = +1\, for\; positive\; samples$$

$$y_i = -1\, for\; negative\; samples$$

Finally, we have our constraints and have defined the gutter in Equation 12.

$$y_i(\vec{x}_i\vec{w} + b) \geq 1$$

$$y_i(\vec{x}_i\vec{w}_i + b) - 1 \qquad \textbf{Eq. 12}$$

$$where\; y_i(\vec{x}_i\vec{w} + b) - 1 = 0\; is\; in\; the\; gutter$$

Now, we want to explain the difference between the gutters as seen in Figure 11. This can be done with Equation 13.

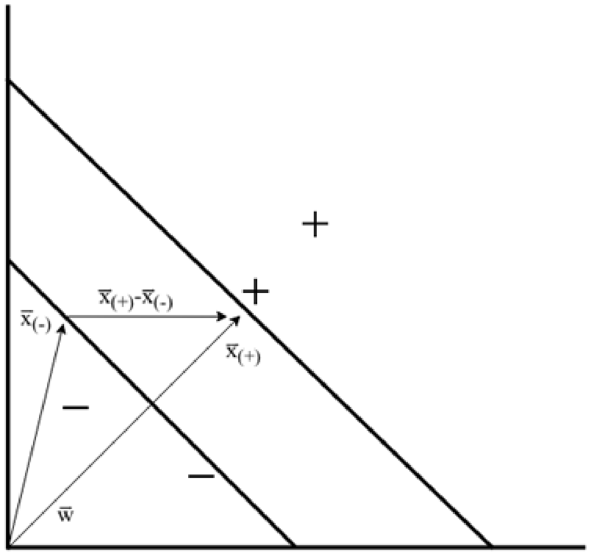

Figure 11: Finding the width of the street

$$width = \left(\vec{x}_{(+)} - \vec{x}_{(-)}\right) \cdot \frac{\vec{w}}{\|w\|} \qquad \textbf{Eq. 13}$$

Now the two gutters are defined and we have determined the width of the street.

$$\begin{aligned} \vec{w} \cdot \vec{x}_{(+)} &= 1 - b \\ \vec{w} \cdot \left(-\vec{x}_{(-)}\right) &= 1 + b \end{aligned} \qquad \textbf{Eq. 14}$$

This is illustrated by our gutter constraint and a reiteration of Equation 14.

$$y_i(\vec{x}_i\vec{w} + b) - 1 = 0$$

$$\vec{x}_i\vec{w} = 1 - b$$

$$\vec{x}_i(-\vec{w}) = 1 + b$$

$$y_i = 1$$

Therefore using,

$$y_i(\vec{x}_i\vec{w} + b) - 1 = 0$$

$$\frac{\vec{x}_{(+)} - \vec{x}_{(-)}}{1} \cdot \frac{\vec{w}}{\|w\|} = \frac{2}{\|w\|}$$

The width of the street is $\left(\frac{2}{\|w\|}\right)$; so we maximize this finding the widest part of the street.

$$\max\left(\frac{2}{\|w\|}\right)\sim\max\left(\frac{1}{\|w\|}\right)\sim\min(\|w\|)\sim\min\frac{1}{2}\|w\|^2$$

Then utilize a LaGrangian to find the maxima.

$$\mathcal{L}=\frac{1}{2}\|\vec{w}\|^2-\sum\nolimits_{\alpha_i}[y_i(\vec{x}_i\vec{w}+b)-1]$$

$$\mathcal{L}=\frac{1}{2}\left(\sum\nolimits_{\alpha_i}y_i\vec{x}_i\right)\cdot\left(\sum\nolimits_{\alpha_j}y_j\vec{x}_j\right)-\left(\sum\nolimits_{\alpha_i}y_ix_i\right)\cdot\left(\sum\nolimits_{\alpha_j}y_jx_j\right)-\left(\sum\nolimits_{\alpha_i}y_ib\right)+\left(\sum\alpha_i\right)$$

$$\mathcal{L}=\sum\alpha_i-\tfrac{1}{2}\sum\sum\alpha_i\alpha_jy_iy_jx_i\cdot x_j$$ **Eq. 15**

Going back to the decision rule illustrates that the maximization depends on the sample vectors. The simplicity of this optimization is that it depends on simple dot products of pairs of samples.

$$\sum\alpha_iy_i\vec{x}_i\cdot\vec{u}+b\geq 0\ \ when\,positive$$ **Eq. 16**

Thus, one can see the decision rule only depends on the dot product of those sample vectors and the unknown. This means there is a total dependence of all the math on the dot products. However, problems arise when the set is not linearly separable. This problem is later solved by Vapnik when he realized the transformation relies on the dot products and can be transformed into any space using a solution known as the kernel trick.

Lastly, if we call a transformation $\phi(\vec{x})$, all we need are the dot products $\phi(x_i)\cdot\ \phi(y_i)$ to find the maximum. Where K is a kernel function:

$$K(x_i,x_j)=\phi(x_i)\cdot\phi(x_j)$$ **Eq. 17**

One of the most frequently used kernel functions, the Radial Basis Function (RBF):

$$K(x_i, x_j) = e^{-\left(\frac{\|x_i - x_j\|}{\sigma}\right)} \quad \textbf{Eq. 18}$$

Introduction to Support Vector Machines Regression

We can see how Support Vector Machines can classify groups of data together and define separation by using a hyperplane. Support Vector Machines operate using a model which depends only on a subset of the training data, because the cost function for building the model does not care about training points which lie beyond the margin. The model for Support Vector Machines-Regression (SVR) also only uses a subset of the training data, because the cost function for building the model ignores any training data close to the model prediction. Using the same model as SVM but instead of maximizing the width of the street we now minimize it. This simply means solving the problem in respect to minimizing the width of the street:

In order to do this, let's consider a set of training points, $\{(\boldsymbol{x}_1, z_1), \ldots, (\boldsymbol{x}_l, z_l)\}$, where $\boldsymbol{x}_i \in \mathbb{R}^n$ is a feature vector and $z_i \in \mathbb{R}^1$ is the target output. Under given parameters $C > 0$ and $\epsilon > 0$, the standard form of support vector regression (Vapnik 1998)

$$\min_{w,b,\xi,\xi^*} \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{l}\xi_i + C\sum_{i=1}^{l}\xi_i^*$$

$$\text{subject to } \boldsymbol{w}^T\phi(\boldsymbol{x}_i) + b - z_i \le \epsilon + \xi_i,$$

$$\boldsymbol{z}_i - \boldsymbol{w}^T\phi(\boldsymbol{x}_i) - b \le \epsilon + \xi_i^*,$$

$$\xi_i, \xi_i^* \ge 0, i = 1, \ldots, l.$$

The dual problem is defined in LIBSVM (Chang 2011)

$$\min_{\alpha,\alpha^*} \; \frac{1}{2}(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)^T Q(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) + \epsilon \sum_{i=1}^{l} (\boldsymbol{\alpha} + \boldsymbol{\alpha}^*) + \sum_{i=1}^{l} z_i(\boldsymbol{\alpha_i} - \boldsymbol{\alpha_i^*})$$

$$\text{subject to } \boldsymbol{e}^T(\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) = 0,$$

$$0 \leq \alpha_i, \alpha_i^* \leq C, i = 1, \ldots, l,$$

$$\text{where } Q_{ij} = K(\boldsymbol{x}_i, \boldsymbol{x}_j) \equiv \; \phi(\boldsymbol{x}_i)^T \phi(\boldsymbol{x}_j)$$

After solving this problem, the approximate function is

$$\sum_{i=1}^{l} z_i(-\boldsymbol{\alpha_i} + \boldsymbol{\alpha_i^*})\, K(\boldsymbol{x}_i, \boldsymbol{x}) + b \qquad \textbf{Eq. 19}$$

Methods of Analysis

In this study, we use LIBSVM, a model written for use with Matlab to perform SVR-∈ on a dataset which was previously analyzed by transformation using Singular Value Decomposition (SVD). This experiment was designed mainly to examine how a statistical learning method would compare with a matrix factorization method. As in the case of the previous study, we examined how well each measured sensory attribute (n=20) could be correctly predicted by concentrations of ethanol (n=6), tannin (n=3), and fructose (n=2). First, two matrices were imported from the previous study into Matlab version 2015a (Mathworks, Natick, MA). One matrix named **A**, contained attribute intensity ratings (864x20). A second matrix, named **C**, contained concentrations of ethanol, tannin, fructose, and a column of ones (864x4) which was utilized as a bias constant in the SVD study previously performed. These matrices had individual panelist data extracted from them which was scaled from 0 to 1. Each panelist dataset underwent the same training and testing procedure which included a 72-fold cross validation and parameter search which optimized the model by minimizing error.

To begin, we scaled matrices **A** and **C** from zero to one. Then, we trained the SVR on each of the attribute ratings provided by the panelists. For every attribute there was a corresponding instance matrix comprised of varying concentrations of ethanol, tannin, and fructose which followed the rules of randomized block design in human sensory testing. Our model was trained using the svmtrain function in Matlab and our code can be found in the Appendix.

The training was performed on each of the panelist's datasets and the model was then tested for accuracy. In this way, we trained on the entire set of data then tested each panelist measurement with our model using the function svmpredict (Appendix). We tested each measurement giving us an output of a predicted value from the model and the measured value which was already provided by the actual panelist measurement giving us a total of 240 predicted values. We calculated our correlation coefficient (R^2) by taking the square of our dot products (Equation 20)

$$R^2 = 1 - \left[\left(\frac{\hat{y}}{\|\hat{y}\|}\right) \cdot \left(\frac{y}{\|y\|}\right)\right]^2 \qquad \textbf{Eq. 20}$$

We computed the correlation coefficient of our model in conjunction with a parameter optimization of the SVR-ϵ**.** The parameter optimization was performed over a range published by the authors of LIBSVM. The parameters we optimize are explicit to the radial basis function kernel and include C, γ, and ϵ**.** These data dependent kernel parameters are tuned using grid search methodology outlined in the supporting documentation for LIBSVM and through using a series of program loops which we embedded into our code (Appendix)(Chang 2011).

Results and Discussion

This experiment was designed to test the hypothesis that each individual's function which converts components in wine into sensory intensities is better predicted by statistical learning algorithm (SVR) over a matrix factorization method (SVD). Our results indicate that while the parameter search and optimization of SVR appears promising and could be better tuned, the SVR model did not outperform our SVD model. During training we noticed early convergence of the model to small values of γ, and ϵ which might indicate an under fitting model. Continued progress on model optimization will most certainly improve our results. Our SVR model proved to be less accurate than our SVD model (Figure 12).

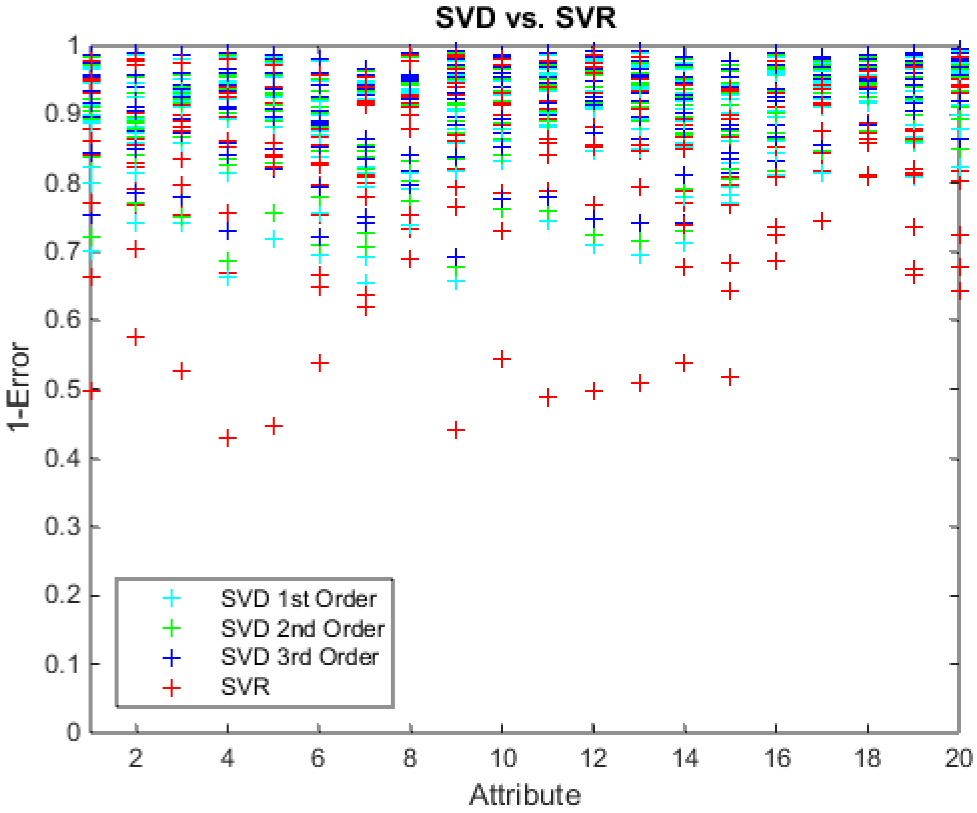

Figure 12: SVD vs. SVR incidence of error

We generally see the SVR does not work as well as the SVD. When we look at each panelist individually, we see an overall underfitting of the SVR model (Figure 13).

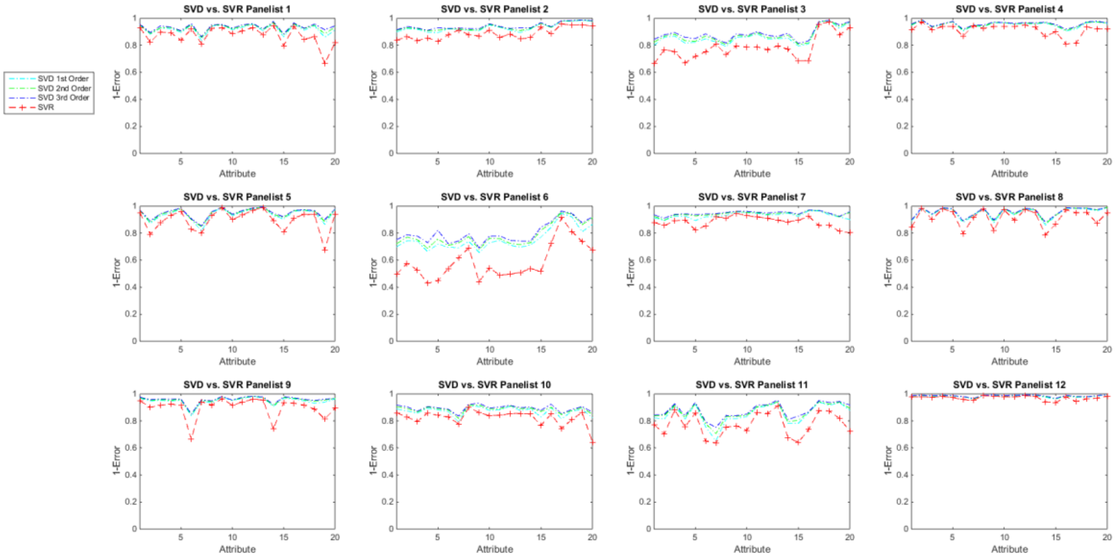

Figure 13: SVD vs. SVR for each panelist

Conclusions

This experiment was designed to test the ability of two regression models in the predicting of wine aroma, flavors, and mouthfeel. While we found the SVD model outperforms the SVR model, both models perform reasonably well in most cases. Generally speaking, the models can accurately predict aromas, flavors, and mouthfeel intensities as a direct result of concentrations of ethanol, tannin, and fructose over 80% of the time. On some occasions we see outlying panelists. These panelists behave erratically compared to the panel as a whole or they exhibit a bias in their responses. SVD and SVR models illustrated difficulty in predicting the responses of these panelists. While SVR did not outperform SVD in this particular study, the promise of SVR is great and relies on a more in depth parameter study to achieve optimal results.

CHAPTER V

CONCLUSIONS AND RECOMMENDATIONS

The overall research objective was to examine the predictability of perceived sensory response based on the concentrations of ethanol, tannin, and fructose. We used two methods commonly found in statistics and statistical learning theory to illustrate how we can predict the panelist response in the present data set based on the composition of the model wine. SVD of aromas vs. flavors illustrated we could predict one from the other and vice versa over 90% of the time. This illustrates high correlation among the shared receptors in the nose and the mouth from sensory feedback we received from our trained panel as intensity ratings. For this data set, we also saw the presence of three major outlying panelists in our SVD study. These were panelists three, six, and eleven. We noticed a dramatic increase in the predictability of the system by removing the outlying panelists. However, these results are not reported due to a decision to keep the outliers as contributors to the group but labelling them as such for additional study and closer observation. The SVM portion of the experiment revealed that the outliers gave bad training data to the machine and therefore may have reduced the machine's ability to predict taste, smell, and mouthfeel. Our inclusion of the outliers was an important part of this study because we feel the outliers are a necessity to report in our observations. The machine learning algorithm did not perform as well as the non-linear SVD but did give relatively high predictive power in some cases. We suspect more data will strengthen the machine learning method and the applications for machine learning in sensory science are still considered in their infantile stages. Possible applications for the SVD and the SVM include smartphone app development, implementing the models to sensor array development and smarter sensor technologies, winemaking blending trials, consumer preference models, and identifying outlying nodes. This data and the predictive models herein provide valuable insight into the future of artificial intelligence and currently has the capacity to make an artificially intelligent sensor array system become more human.

BIBLIOGRAPHY

Amerine, MA, Berg, HW; Cruess, WV (1972). The Technology of Wine Making. Westport Ct., Avi Publishing Co. .

Amerine, MA, Ough, CS; Gatlander, JF (1972). Recent Advances in Enology. CRC critical reviews in Food Technology **2**(4): 407-515.

Amerine, MA, Roessler, EB; Filipello, F (1959). "Modern Sensory Methods of Evaluating Wine." Hilgardia **28**(18): 477-561.

Aznar, M, Lopez, R, Cacho, J; Ferreira, V (2003). Prediction of aged red wine aroma properties from aroma chemical composition. Partial least squares regression models. Journal of Agriculture Food Chemistry **51**: 2700-2707.

Bakalar, N (2012). Partners in flavour. Nature **486**.

Baker, GA, Ough, CS; Amerine, MA (1966). "Sensory Scores and Analytical Data for Dry White Wine and Dry Red Wines as Bases for Predictions and Groupings." American Journal of Enology and Viticulture **17**(4): 255-264.

Barnett, PN, Blair, N; Gardner, JW (1993). Electronic noses. Principles, applications and outlook. Proceedings of the 15th ASIC Colloque, Montpellier.

Carter, GH, Nagel, CW, Nelson, J, Attalah, M, Johnson, T, Early, R; Clore, WJ (1974). "A Summary of Experimental Testing of Grape Varieties for Wine in Washington." American Journal of Enology and Viticulture **25**(2): 92-98.

Cattell, RB (1966). The scree test for the number of factors. Multiv. Behavior Res **1**.

Cetó, X, Apetrei, C, del Valle, M; Rodríguez-Méndez, ML (2014). "Evaluation of red wines antioxidant capacity by means of a voltammetric e-tongue with an optimized sensor array." Electrochimica Acta **120**: 180-186.

Cetó, X, Capdevila, J, Mínguez, S; del Valle, M (2014). Voltammetric BioElectronic Tongue for the analysis of phenolic compounds in rosé cava wines. Food Research International **55**: 455-461.

Chang, C-CaL, Chih-Jen (2011). "LIBSVM: A library for support vector machines." ACM Transactions on Intelligent Systems and Technology **2**(3): 27:21--27:27.

Chatfield, C; Collins, AJ (1980). Introduction to Multivariate Analysis. London, Chapman and Hall.

Cortez, P, Cerdeira, A, Almeida, F, Matos, T; Reis, J (2009). "Modeling wine preferences by data mining from physicochemical properties." Decision Support Systems **47**(4): 547-553.

Dehlholm, C, Brockhoff, PB; Bredie, WLP (2012). "Confidence ellipses: A variation based on parametric bootstrapping applicable on Multiple Factor Analysis results for rapid graphical evaluation." Food Quality and Preference **26**(2): 278-280.

Dixon, WJ (1974). BMD Biomedical Computer Programs. Berkely and Los Angeles, Ca, University of California Press.

Dravnieks, A, Reilich, HG; Whitfield, J (1973). Classification of corn odor by statistical analysis of gas chromatographic patterns of headspace volatiles. Journal of Food Science. **38**: 34-39.

Duncan, DB (1955). Multiple range and multiple F tests. Biometrics **11**: 1-42.

Ferrier, JG; Block, DE (2001). Neural-network-assisted optimization of wine blending based on sensory analysis. American Journal of Enology and Viticulture **52**(4): 366-395.

Gianturco, MA, Biggers, RE; Ridling, BH (1974). Seasonal variations in the composition of the volatile constituents of black tea. A numerical approach to the correlation between composition and quality of tea aroma. Journal of Agriculture Food Chemistry **22**: 758-764.

Gnanadesikan, R; Wilk, MB (1969). Data analytic methods in multivariate statistical analysis. In "Multivariate Analysis II". London, Academic Press.

Gómez-Meire, S, Campos, C, Falqué, E, Díaz, F; Fdez-Riverola, F (2014). Assuring the authenticity of northwest Spain white wine varieties using machine learning techniques. Food Research International **60**: 230-240.

Green, PE (1978). Analyzing Multivariate Data. Hinsdale, Il, Dryden Press.

Guignard, JX; Cliff, M (1987). Descriptive analysis of Pinot noir from Carneros, Napa, and Sonoma. American Journal of Enology and Viticulture. **38**(4).

Heymann, H; Noble, AC (1987). Descriptive Analysis of Commercial Cabernet Sauvignon Wines from California. American Journal of Enology and Viticulture **38**(1): 41-44.

Heymann, H; Noble, AC (1989). Comparison of Canonical Variate and Principal Component Analyses of Wine Descriptive Analysis Data. Journal of Food Science **54**(5): 1355-1358.

Hoff, JT, Helbert, JR; Chicoye, E (1975). Classification of lager beers by computer analysis of volatile profiles. MBAA Tech Quarterly **12**: 209-213.

Hosu, A, Cristea, VM; Cimpoiu, C (2014). Analysis of total phenolic, flavonoids, anthocyanins and tannins content in Romanian red wines: prediction of antioxidant activities and classification of wines using artificial neural networks. Food Chemistry **150**: 113-118.

Kaiser, HF (1960). "The application of electronic computers to factor analysis." Educ. Psychol. Measurement **20**.

Kohl, D (1996). Semiconductor and calorimetric sensor devices and arrays. Handbook of biosensors and electronic noses: Medicine, food & the environment. E. Kress-Rogers. Boca Raton FL, CRC Press Inc.

Kwan, WO; Kowalski, BR (1980). Correlation of objective chemical measurements and subjective sensory evaluations: Wines of Vitia vinifera variety 'Pinot noir' from France and the United States. Analytical Chimica Acta: 122:216.

Lawless, HT (2013). Quantitative Sensory Analysis : Psychophysics, Models and Intelligent Design. West Sussex, UK, John Wiley & Sons.

Lawless, HT; Heymann, H (1998). Sensory evaluation of food, principles and practice. New York, USA, Chapman and Hall.

Leifu Gao, XL (2009). A resilient particle swarm optimization algorithm based on chaos and applying it to optimize the fermentation process. International Journal Of Information and Systems Sciences **5**(3-4): 380-391.

Mardia, KVK, J.T.; Bibby, JM (1979). Multivariate Analysis. London, Academic Press.

Matlab (2015). Matlab. Mathworks. Natick Massachusettes, Mathworks. **2015**.

Mitrovics, J, Ulmer, H, Noetzel, G, Weimar, U; Goepel, W (1997). Design of a hybrid modular sensor system for gas and odor analysis. Proceedings of the transducers 97 conference, Chicago.

Moll, M, Flayeax, R, That, V; Noel, JPEBI- (1974). Relations entre les paramitres physico-chemiques de la biere et las resultats de de-gustations. European Brewers International **9**: 328-333.

Noble, AC (1975). Instrumental analysis of the sensory properties of food. Food Technology **29**: 56-60.

Noble, AC; Ebeler, SE (2002). Use of multivariate statistics in understanding wine flavor. Food Reviews International **18**(1): 1-21.

Noble, AC; Shannon, M (1987). Profiling Zinfandel wines by chemical and sensory analyses. American Journal of Enology and Viticulture **38**(1).

Noble, AC, Williams, A-A; Langron, SP (1984). Descriptive analysis and quality ratings of 1976 wines from four Bordeaux communes. Journal of Science Food Agriculture **35**.

O'Mahony, M (2013). The Tetrad Test: Looking Back, Looking Forward. Journal of Sensory Studies **28**(4): 259-263.

Ough, CS; Baker, GA (1961). Small panel sensory evaluations of wines by scoring. Hilgardia **30**(19): 587-619.

Ough, CS; Cordner, CW (1978). Prediction Of Panel Preference For Zinfandel Wine From Analytical Data: Using Difference In Crop Level To Affect Must, Wine, And Headspace Composition. American Journal of Enology and Viticulture **29**(4).

Penza, M; Cassano, G (2004). Chemometric characterization of Italian wines by thin-film multisensors array and artificial neural networks. Food Chemistry **86**(2): 283-296.

Schaefer, J, Tas, AC, VeliBek, J, Maarse, H, ten Noever de Brauw, MC; Slump, P (1983). Application of pattern recognition techniques in the differentiation of wines. London, Academic Press.

Scheffé, H (1952). An analysis of variance for paired comparisons. Journal of American Statistics Association **47**: 381-400.

Selih, VS, Sala, M; Drgan, V (2014). Multi-element analysis of wines by ICP-MS and ICP-OES and their classification according to geographical origin in Slovenia. Food Chem **153**: 414-423.

Shanmuganathan, S; Whalley, J (2013). Pixel clustering in spatial data mining; an example study with kumeu wine region in New Zealand. 20th International Congress on Modelling and Simulation, Adelaide, Australia.

Singleton, VL, Sieberhagen, HA, de Wet, P; van Wyk, CJ (1975). Composition and sensory qualities of wines prepared from white grapes by fermentation with and without grape solids. American Journal of Enology and Viticulture **26**(2): 62-69.

Stephan, A, Bucking, M; Steinhart, H (2000). Novel analytical tools for food flavours. Food Research International **33**: 199-209.

Stone, H, Bleibaum, R; Thomas, H (2012). Sensory Evaluation Practices San Diego, CA, Elsevier Academic Press

Sun, L-X, Danzer, K; Thiel, G (1997). "Classification of wine samples by means of artificial neural networks and discrimination analytical methods." Fresnius J Anal Chem **359**: 143-149.

Tao, Y, Wu, D, Zhang, QA; Sun, DW (2014). "Ultrasound-assisted extraction of phenolics from wine lees: modeling, optimization and stability of extracts during storage." Ultrason Sonochem **21**(2): 706-715.

Tatsuoka, M (1971). Multivariate Analysis. Techniques for Educational and Psychological Research. New York, John Wiley.

Vapnik, V (1998). Statistical learning theory, Wiley New York.

Veeramachaneni, K, Vladislavleva, E; O'Reilly, U-M (2012). "Knowledge mining sensory evaluation data: genetic programming, statistical techniques, and swarm optimization." Genetic Programming and Evolvable Machines **13**(1): 103-133.

Villamor, RR (2012). The Impact of Wine Components on the Chemical and Sensory Properties of Wines. School of Food Science, Washington State University. **Doctor of Philosophy:** 123.

Villamor, RR, Evans, MA, Mattinson, DS; Ross, CF (2013). "Effects of ethanol, tannin and fructose on the headspace concentration and potential sensory significance of odorants in a model wine." Food Research International **50**(1): 38-45.

Vlassides, S, Ferrier, JG; Block, DE (2001). "Using historical data for bioprocess optimization: modelling wine characteristics using artificial neural networks and archived process information." Biotechnol Bioeng **73**(1): 55-67.

Wang, R, Zeng, W; Ming, J (2014). "Discrimination of the White Wine Based on Sparse Principal Component Analysis and Support Vector Machine." **277**: 695-702.

Winston, PH (2010). Lecture 16: Learning: Support Vector Machines. Electrical Engineering and Computer Science: Artificial Intelligence: Lecture Videos. Massachusetts Institute of Technology, MIT OpenCourseWare.

APPENDIX

MATLAB CODE

```
    clear
    %Load Set
    load('indep.mat')%ethanol, tannin, fructose (864x3)
    load('dep.mat')%crmlckdroma, rthmshroomroma, flrlroseroma, fruitroma,
            %grnvgtlroma, spiceclveroma, slfrchemroma, woodmedroma,
            %crmlckdflvr, rthmshroomflvr, flrlroseflvr, fruitlflvr,
            %grnvgtlflvr, spiceclveflvr, slfrchemflvr, woodmedflvr,
            %bittrnssmthfl, drymthfl, heatmthfl, sournessmthfl (864x20)
    x=ones(size(dep,1),1);
    %%%%%%%%%%%%%%%%%%%%%%%%%%%
    for MO = 1:3
    if (MO == 1)
      indepx=[indep,x];


    elseif (MO == 2)
    indepx=[indep(:,1).^2,indep(:,2).^2,indep(:,3).^2,indep(:,1).*indep(:,2),indep(:,1).*in
dep(:,3),indep(:,2).*indep(:,3),indep(:,1),indep(:,2),indep(:,3),x];


    elseif (MO == 3)
    indepx=[indep(:,1).^3,indep(:,2).^3,indep(:,3).^3,indep(:,1).^2.*indep(:,2),indep(:,1).
^2.*indep(:,3), indep(:,2).^2.*indep(:,1),
indep(:,2).^2.*indep(:,3),indep(:,3).^2.*indep(:,1),indep(:,3).^2.*indep(:,2),
indep(:,1).*indep(:,2).*indep(:,3), indep(:,1).^2, indep(:,2).^2, indep(:,3).^2,
indep(:,1).*indep(:,2), indep(:,1).*indep(:,3), indep(:,2).*indep(:,3), indep(:,1), indep(:,2),
indep(:,3), x];
    end;
    %%%%%%%%%%%%%%%%%%%%%%%%%%%
```

```
for r = 1:72:864; %location of first sample for each panelist
   pnlstindep = indepx(r:r+71,:); %get panelist data from indep and dep
   pnlstdep = dep(r:r+71,:);
   pnlstvarindep{(r+71)/72}=pnlstindep; %store pnlst info as vector in an array
   pnlstvardep{(r+71)/72}=pnlstdep; %store pnlst info as vector in a matrix
end
%Compute SVD on each panelist, then array missing one, two, three, and four
%and store results.
for A=1:1:12;
     [u,s,v]          = svd(pnlstvarindep{1,A});
     [numSamp,dimInp]  = size(s);
%%%  We need not use singular values that are close to zero
     cut_off  = diag(s) > max(diag(s))*10^(-10);
    % a        = diag(cut_off.*(1./diag(s)),dimInp,numSamp);
     a    = [diag(cut_off.*(1./diag(s))),zeros(dimInp, numSamp-dimInp)];
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
%%%%%%%%%%%%%%%%%%%%%%%%%
   for AA=1:1:20, pnlstvardep{1,A}(:,AA);
    %Regression coefficients for each attribute of each panelist in array
     beta{A}{AA}=v*a*u'*pnlstvardep{1,A}(:,AA);
    % error{A}{AA}=pnlstvarindep{A}*beta{A}{AA}-pnlstvardep{A}(:,AA);
    %
error{A}{AA}=dot(pnlstvarindep{A}*beta{A}{AA}/norm(pnlstvarindep{A}*beta{A}{AA}),
pnlstvardep{A}(:,AA)/norm(pnlstvardep{A}(:,AA)));
     rsquare(MO,A,AA)=1-(norm(error{A}{AA})/norm(pnlstvardep{A}(:,AA)))^2;
     %plot3(A,AA,rsquare(MO,A,AA),'r+');
   end
    mean_rsquare(MO,A) = mean(rsquare(MO,A,:));
```

```
end
end % MO loop



%%%%%% Plot %%%%%%%%%%%%%%%%%%

[XX,YY] = meshgrid([1:20],[1:12]);

figure;
hold on;
plot3(XX,YY,reshape(rsquare(1,:,:),12,20),'r+');
plot3(XX,YY,reshape(rsquare(2,:,:),12,20),'b+');
plot3(XX,YY,reshape(rsquare(3,:,:),12,20),'g+');
hold off;
max2_1 = max(max(rsquare(2,:,:)-rsquare(1,:,:)))
max3_2 = max(max(rsquare(3,:,:)-rsquare(2,:,:)))
max3_1 = max(max(rsquare(3,:,:)-rsquare(1,:,:)))

min2_1 = min(min(rsquare(2,:,:)-rsquare(1,:,:)))
min3_2 = min(min(rsquare(3,:,:)-rsquare(2,:,:)))
min3_1 = min(min(rsquare(3,:,:)-rsquare(1,:,:)))

avg2_1 = sum(sum(rsquare(2,:,:)-rsquare(1,:,:)))/240
avg3_2 = sum(sum(rsquare(3,:,:)-rsquare(2,:,:)))/240
avg3_1 = sum(sum(rsquare(3,:,:)-rsquare(1,:,:)))/240

%%% SVM 1/17/2016%%%%%%%%%%%%%%%%%%%
clear
%%% Initial Stuff
```

```
load indep.mat
load dep.mat
x=ones(size(dep,1),1);
indepx=[indep,x];

%% scaling
% indep_scaled
    [maxV, I] = max(indepx);
    [minV, I] = min(indepx);
    [R, C] = size(indepx);
    scaled = (indepx-ones(R, 1)*minV).*(ones(R, 1)*((1-0)*ones(1, C)./(maxV-minV))) +0;

    for i = 1:size(indepx, 2)
        if (all(isnan(scaled(:, i))))
            scaled(:, i) = 0;
        end
    end
    indep_scaled = scaled;
% dep_scaled
    [maxV, I] = max(dep);
    [minV, I] = min(dep);
    [R, C] = size(dep);
    scaled = (dep-ones(R, 1)*minV).*(ones(R, 1)*((1-0)*ones(1, C)./(maxV-minV))) +0;

    for i = 1:size(dep, 2)
        if (all(isnan(scaled(:, i))))
            scaled(:, i) = 0;
```

```
      end
   end
   dep_scaled = scaled;



%% Parameter selection
param.s = 3;                                    % should do epsilon #3 and nu SVR #4 both
param.t = 2;                                    % RBF kernel
param.C = 1;                  % C (could do more param search)
param.gset = 2.^[-2:7];                         % Range of the gamma parameter
param.eset = [0:5];         % Range of epsilon parameter
%param.nuset = [0:.05:1];        % Range of nu parameter

for r = 1:72:864; %location of first sample for each panelist
          pnlstindep = indep_scaled(r:r+71,:); %get panelist data from indep and dep
          pnlstdep = dep_scaled(r:r+71,:);
          pnlstvarindep{(r+71)/72}=pnlstindep; %store pnlst info as vector in an array
          pnlstvardep{(r+71)/72}=pnlstdep; %store pnlst info as vector in a matrix
end
for A = 1:1:12
   for AA = 1:1:20
   trn.X=[pnlstvarindep{A}(:,:)]; %EtOH, Tannin, Fructose (Instance Matrix)
   trn.Y=[pnlstvardep{A}(:,AA)]; %Attribute (Label Vector)
    MSE = zeros(length(param.gset), length(param.eset)); %preallocation
    kevin_error = zeros(length(param.gset), length(param.eset)); %preallocation
      for j = 1:length(param.gset);
         param.g = param.gset(j);
```

```
            for k = 1:length(param.eset);
                y_hat = zeros(72,1); %preallocation
                param.e = param.eset(k);
                param.libsvm = ['-s ', num2str(param.s), ' -t ', num2str(param.t), ' -c ',
num2str(param.C), ' -g ', num2str(param.g), ' -p ', num2str(param.e)];
                for i = 1:72;
                    tst.X = trn.X([i]); %test instance
                    tst.Y = trn.Y([i]); %test label
                    array.X = trn.X([1:i-1, i+1:72],:); %training instance matrix
                    array.Y = trn.Y([1:i-1, i+1:72],:); %training label vector
                    model = svmtrain(array.Y,array.X,param.libsvm); %svmtrain
                    [y_hat(i),acc,prediction] = svmpredict(tst.Y,tst.X,model); %svmpredict
                    MSE(j,k) = MSE(j,k) + abs((y_hat(i)-tst.Y).^2); %Mean Squared Error
                    kevin_error(j,k)=dot(y_hat/norm(y_hat),trn.Y/norm(trn.Y)); %Kevin
requested error
                end
            end
        end
    MSE = MSE ./ 71; %Degrees of freedom
    [v1, i1] = max(kevin_error); %Maximum Error
    [v2, i2] = min(v1); %Minimum Error
    optparam = param;
    optparam.g = param.gset( i1(i2) ); %Optimized Parameter Gamma
    optparam.e = param.eset(i2); %Optimized parameter Epsilon
    optparam.libsvm = ['-s ', num2str(optparam.s), ' -t ', num2str(optparam.t), ...
                ' -c ', num2str(optparam.C), ' -g ', num2str(optparam.g), ...
                ' -p ', num2str(optparam.e)];
        %% Optimized output
    optimal_kevin_error{A}{AA} = kevin_error(i1(i2),i2); %Optimal Error captured
and reported
```

```
%% Run model with optimized parameters
model = svmtrain(trn.Y, trn.X, optparam.libsvm); %Using all data for training
[y_hat, Acc, projection] = svmpredict(trn.Y, trn.X, model); %Testing on all, (note Trn.Y & Trn.X & model)
predvstrue{A,AA} = [y_hat,trn.Y]; %Optimized Prediction vs True
    end
end


clear
load 'optimal_kevin_error.mat' %Error from SVR
load 'rsquare.mat' %rsquare from SVD
load 'predvstrue.mat' %predvstrue from SVR


%% SVD 1st ORDER VS. SVR concatenation
for A=1:12
    optimal_kevin_error{A}; %Error from SVR
    form2=permute(rsquare(:,A,:),[3,1,2]); %Error from SVD
    form1=cell2mat(optimal_kevin_error{A});
    scores{A}=[form2,form1'];%SVD results are first three columns, SVR is last one
end
% SVD all Orders vs SVR Plots
figure
for i=1:12
  subplot(3,4,i)
w=scores{i}(:,1); % SVD 1st Order cyan
x=scores{i}(:,2); % SVD 2nd Order green
y=scores{i}(:,3); % SVD 3rd Order blue
z=scores{i}(:,4).^2; % SVR red; % SVR red
plot(w,'-.c')
```

```
hold on

plot(x,'-.g')

plot(y,'-.b')

plot(z,'--+r')
```

%%% The form of this code is provided illustrating the creation of the SVD and SVM models and does not include the datamining involved or the constructs designed to manipulate these two functions. ddycus@yahoodotcom<END>