STAT 5330 Project 2: Wine Quality Assessment

Introduction:

Source: *This dataset is publicly available for research. The details are described in [Cortez et al., 2009]. P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis. Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4): 547-553. ISSN: 0167-9236.*

- **Data Overview :**

  The data set is on white variants of the Portuguese "Vinho Verde" wine with 4898 observations and are collected on 12 different properties of the wines one of which is the response variable, Quality, based on sensory data, and the rest 11 variables are on chemical properties of the wines including density, acidity etc. All chemical properties of wines are continuous variables. Quality is an ordinal variable with possible ranking from 1 (worst) to 10 (best). Each variety of wine is tasted by three tasters and the final rank assigned is the median rank given by the tasters.

- **Background Information:**

  Nowadays, wine is increasingly enjoyed by a great range of markets. Therefore, wine certification and quality assessment become key elements in the industry within this context. Wine certification is often assessed by physicochemical and sensory tests which are laboratory-based and takes into account factors like acidity, pH level, presence of sugar and other chemical properties, while sensory tests rely mainly on human experts, opinion among whom may have a high degree of variability.

- **Variables Information and Descriptions:**

  Chemical properties of wines (continuous variables):

  1 - fixed acidity (tartaric acid - g / dm^3): most acids involved with wine or fixed or nonvolatile

2 - volatile acidity (acetic acid - g / dm^3): the amount of acetic acid in wine, which at too high of levels can lead to an unpleasant, vinegar taste

3 - citric acid (g / dm^3): found in small quantities, citric acid can add flavor

4 - residual sugar (g / dm^3): the amount of sugar remaining after fermentation stops

5 - chlorides (sodium chloride - g / dm^3): the amount of salt in the wine

6 - free sulfur dioxide (mg / dm^3): the free form of SO2 exists in equilibrium between molecular SO2 (as a dissolved gas) and bisulfite ion; it prevents microbial growth and the oxidation of wine

7 - total sulfur dioxide (mg / dm^3): amount of free and bound forms of S02

8 - density (g / cm^3): the density of water is close to that of water depending on the percent alcohol and sugar content

9 - pH: describes how acidic or basic a wine is on a scale from 0 (very acidic) to 14 (very basic)

10 - sulphates (potassium sulphate - g / dm3): a wine additive which can contribute to sulfur dioxide gas (S02) levels

11 - alcohol (% by volume): the percent alcohol content of the wine

Output variable (based on sensory data):

12 - quality (score between 0 and 10)

Summary Statistics:

	Mean	Median	Min	Max	Max - min	Standard deviation
Fixed acidity	6.855	6.8	3.8	14.2	10.4	0.844
Volatile acidity	0.278	0.26	0.08	1.1	1.02	0.101
Citric acid	0.334	0.32	0	1.66	1.66	0.121
Residual sugar	6.391	5.2	0.6	65.8	65.2	5.072
Chlorides	0.046	0.043	0.009	0.346	0.337	0.022
Free sulfur dioxide	35.308	34	2	289	287	17.007

Total sulfur dioxide	138.361	134	9	440	431	42.498
Density	0.994	0.99374	0.98711	1.03898	0.05187	0.003
PH	3.188	3.18	2.72	3.82	1.1	0.151
Sulphates	0.49	0.47	0.22	1.08	0.86	0.114
Alcohol	10.514	10.4	8	14.2	6.2	1.231
Quality	5.878	6	3	9	6	0.886

Objectives:

Using analytical approaches, such as multiple regression, Stepwise LM, Shrinkage Methods, Dimension Reduction, SVM, Neural Networks, etc., to interpret how human quality of tasting can be related to the chemical properties of wine and further to make prediction of quality ranking from the chemical properties of the wines. When applying these methods, performance highly depends on a correct variable and model selection. The team will show the impact of the obtained model and prediction errors associated with each method to decide the final best model. Finally, the team is also going to answer the three questions of interest in the discussion part.

Question of Interests:

1. How human quality of tasting is related to the chemical properties of wine? We want to determine which variables best influence the overall quality of a white wine. We consider to use regression model such as best-subset, stepwise methods shrinkage and dimension reduction method to identify which variables in the model are most important.

2. We are also interested in the final best model equation from each regression type we used for question 1. We are going to determine the best model based on the test prediction error. Eventually, we would like to make comparison between best models from different regression types based on prediction accuracy to understand which regression methods give more desirable results in our case.

3. Can we classify good or bad quality of white wine using the 11 properties of white wine? Applying three models, random forest, Neural Networks, and Support Vector Machines, on the dataset to illustrate which performance of model type is better. We will first treat the quality score as a factor variable (0-5 represents bad and 6-9 represents good). We want to compare the models and determine the best model obtained through cross-validation in order to further make predictions. We believe that such model is useful to support wine tasting evaluations and improve wine production.

Methods:

1. **Best subset Selection**

   To perform best subset selection, the team fits a separate least squares regression best subset selection for each possible combination of the 11 predictors in wine data set. The team then looks at all of the resulting models, with the goal of identifying the one that is best**.**

2. **Stepwise Selection**

   There might be some constraints and disadvantages associated with using best subset selection, thus stepwise methods come into play to explore a for more restricted set of models.

   a. Forward:

      Forward stepwise selection begins with a model containing no predictors, and then adds predictors to the model, one-at-a-time, until all of the predictors are in the model.

   b. Backward:

      It begins with the full least squares model containing all 8 predictors in the prostate data set, and then iteratively removes the least useful predictor, one-at-a-time.

   c. Hybrid:

      Hybrid versions of forward and backward stepwise selection works when variables are added to the model sequentially, in analogy to forward selection.

3. **Shrinkage method**

   As an alternative to best subset method, the team can fit a model containing all 11 predictors using a technique that constrains or regularizes the coefficient estimates, or equivalently, that shrinks the coefficient estimates towards zero.

   a. Ridge regression:

      Ridge regression is very similar to least squares, except that coefficients are estimated by minimizing a tuning parameter lambda. The tuning parameter λ serves to control the relative impact of these two terms on the regression coefficient estimates. Therefore, selecting a good value for λ is critical.

b. Lasso regression:

Lasso has the same formulation as ridge regression, the only difference is that lasso coefficient penalty has the effect of forcing some of the coefficient estimates to be exactly equal to zero when the tuning parameter λ is sufficiently large.

4. **Dimension reduction method:**

What differentiate the dimension reduction methods with the methods mentioned before is that dimension reduction first transforms the predictors and then fits a least squares model using the transformed variables, so this technique no longer uses original predictors in the model.

a. Principle Component Regression (PCR) involves constructing principal components regression the first M principal components, and then using these components as the predictors in a linear regression model that is fit using least squares.

b. Partial Least Squared (PLS) is a dimension reduction method, which first identifies a new set of features that are linear combinations of the original predictors, and then fits a linear model via OLS using new features.

5. **Support Vector Machines:**

Support Vector Machine maps the data to a higher dimensional space where the data can be linearly separable. It is resulted from enlarging the feature space in a specific ways using kernels $f(x) = \beta 0+\Sigma\alpha iK(xi, xi'))$ to find a support vector classifier in a higher-dimensional space to separate the data. The support vectors are plotted as crosses and the remaining observations are plotted as circles. Three different kernels were implemented: linear, radial, and polynomial. The team then performed ten-fold cross-validation on the three models of interest.

6. **Neural Networks:**

The team employs Artificial Neural Networks to work with the dataset. The nodes in the ANN combine the inputs by multiplying each input by a weight (to simulate its importance) and summing the products. Then, the output from the combination function (a 1 if the result is greater than some threshold and -1 otherwise) is fed into a transfer

function where the perceptron receives inputs, multiplies them by some weight and produces an output. Once having the output, the team can compare it to a known label and adjust the weights accordingly (the weights usually start off with random initialization values) (Portilla). This process is repeated as many times as needed until the perceptron classifies all training examples correctly. Through backward propagation, the team tunes network parameters to best fit the training dataset.

7. **Random Forest:**

   In the Random Forest approach, a large number of decision trees are created. Every observation is fed into every decision tree. The most common outcome for each observation is used as the final output. A new observation is fed into all the trees and taking a majority vote for each classification mode. The team applied random forest in order to classify the quality of white wine since it's able to deal with unbalanced data.

Results:

1. **Best Subset Selection:**

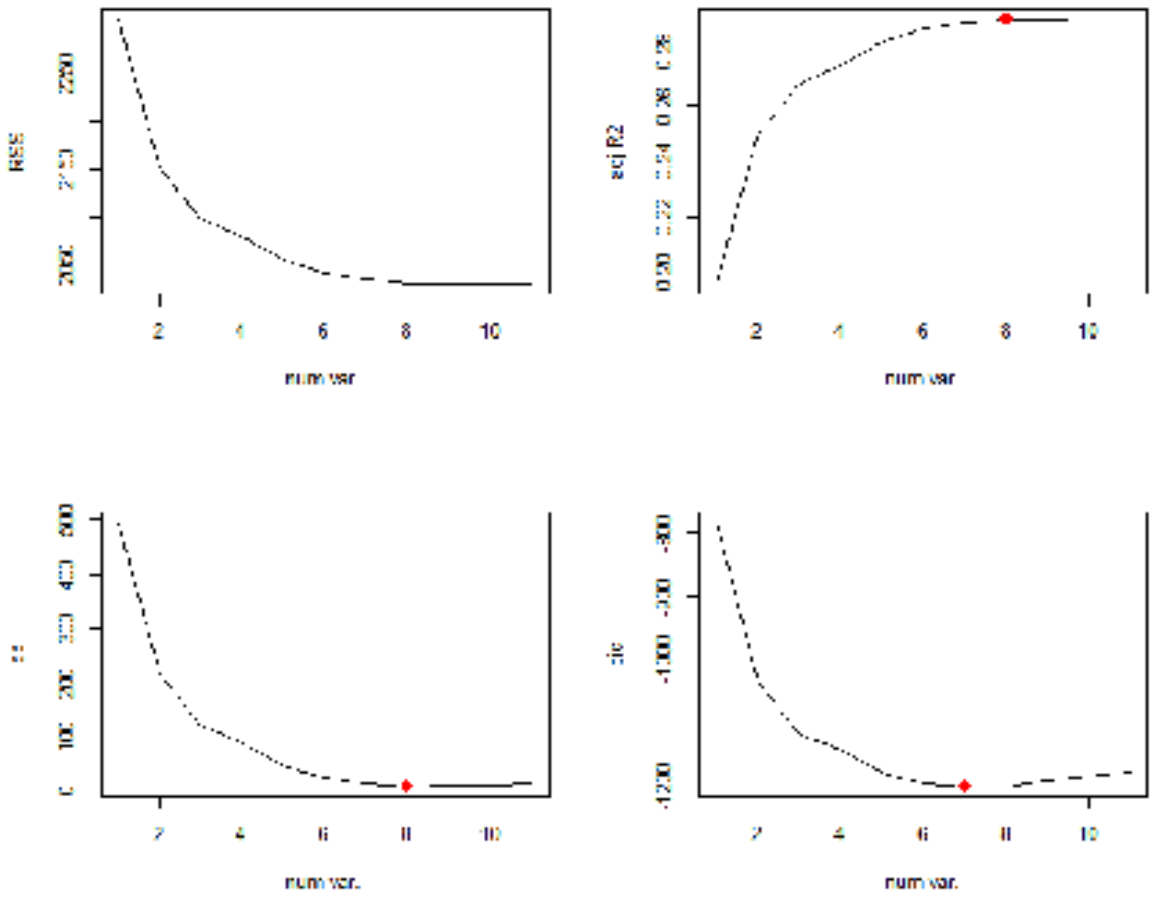

Figure 1: Plot RSS, adjusted R^2, Cp, and BIC for all of the models

Cp selects the **eight-variable model**; BIC estimate of test error shows an increase after seven variables are selected, so BIC suggests a **seven-variable model**. Unlike RSS, Cp and BIC, a large value of adjusted R2 indicates a model with a small test error, so adjusted R^2 chooses the **eight-variable model**.

An alternative to above techniques is that the team can directly estimate the test error then select the model for which the resulting estimated test error is smallest.

	1	2	3	4	5	6
Prediction Error	0.657440	0.626564	0.6134722	0.611017	0.610258	0.60400
	7	8	9	10	11	
Prediction Error	0.5989422	**0.597315**	0.5975462	0.597823	0.597692	

Table 1: Test average prediction error from validation

Final best model on entire dataset (best-subset):

Validation approach leads to the selection of **eight-variable model** which contains :

Quality = (Intercept) +B1 fixed.acidity + B2 volatile.acidity +B3 residual.sugar + B4 free.sulfur.dioxide + B5 density + B6 ph + B7 sulphates +B8 alcohol

where:

Intercept	B1	B2	B3	B4
1.541062e+02	6.810394e-02	-1.888140e+00	8.284724e-02	3.349015e-03
B5	B6	B7	B8	
-1.542913e+02	6.942135e-01	6.285081e-01	1.931628e-01	

Table 2: final model equation from best subset selection on entire data set

2. <u>Stepwise Selection:</u>

Forward, backward and hybrid selections give the team same results. The best model from stepwise selection contain eight variables:

Quality = (Intercept) +B1 fixed.acidity + B2 volatile.acidity +B3 residual.sugar + B4 free.sulfur.dioxide + B5 density + B6 ph + B7 sulphates +B8 alcohol.

This model is exactly the same as the model from best subset selection. The team then computes the prediction error for best model from each stepwise method. Since three stepwise methods yield the same model, the prediction error should be the same. The prediction error from validation gives **0.5973154.** Final model on entire data set would be

the same as what the team gets from best subset selection (Table 2), because the eight variables are the same.

3. <u>Shrinkage:</u>

The team applies Ridge and Lasso regression on the training data set. Unlike least squares, which generates only one set of coefficient estimates, ridge and lasso regression will produce a different set of coefficient estimates, for each value of λ, so selecting a good value for λ is critical. By calculating the test average prediction error for each regression, the team then can choose the best lambda based on prediction error.

	test prediction error. min	Lambda
Ridge Regression	0.5969285	0.03910185
Lasso Regression	0.5973147	0.001222213

Table 3: Test prediction error and corresponding lambda for shrinkage

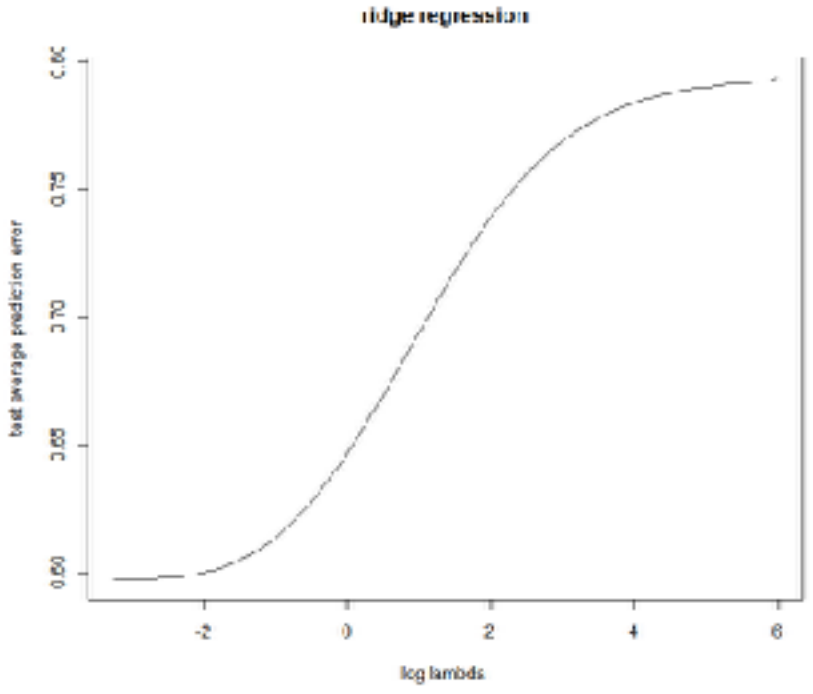

Figure 2: test average prediction error for ridge regression

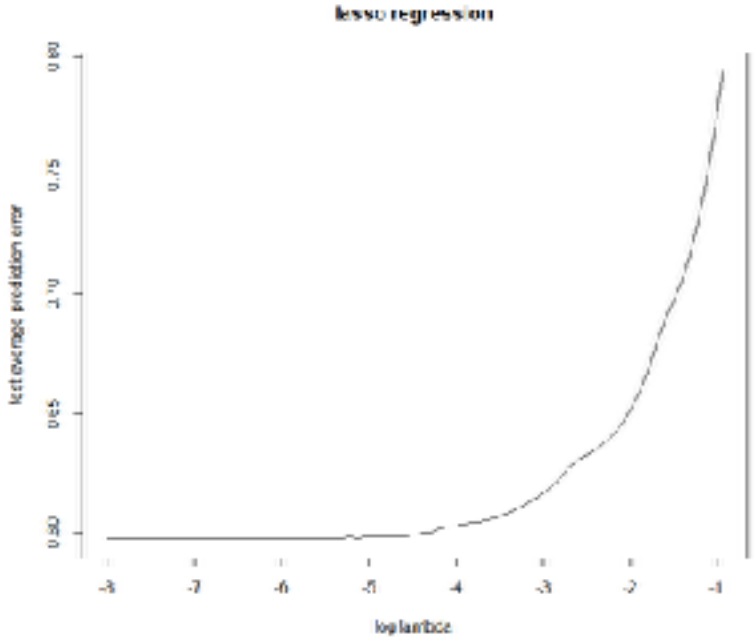

Figure 3: test average prediction error for lasso regression

The final model from ridge regression contains all the 11 variables because ridge-regression won't select variables, the final model on entire data set from ridge regression is:

Quality = (Intercept) +B1 fixed.acidity + B2 volatile.acidity +B3 citric .acid + B4 residual.sugar + B5 chlorides + B6 free.sulfur.dioxide + B7 total.sulfur.dioxide + B8 density + + B9 ph + B10 sulphates +B11 alcohol

where:

Intercept	B1	B2	B3	B4	B5
6.9021e+01	-2.7241e-03	-1.7943e+00	2.2304e-02	4.608e-02	-1.032e+00
B6	B7	B8	B9	B10	B11
4.2284e-03	-6.23671e-04	-6.75301e+01	3.69267e-01	4.9324e-01	3.6926e-01

Lasso regression has the effect of setting the variables exactly equal to zero while ridge does not. However, in the wine data set, the lasso regression method yields a eleven-variable model as well. The coefficients from lasso regression are different from the coefficients from ridge regression even though both method use all the eleven variables.

Intercept	B1	B2	B3	B4	B5
1.3151e+02	4.7718e-02	-1.86998e+00	8.9134e-03	7.380e-02	-3.224e-01
B6	B7	B8	B9	B10	B11
3.6776e-03	-2.6546e-04	-1.3124e+02	6.0266e-01	5.936e-01	2.1360e-01

4. <u>Dimension reduction method:</u>

- **Principle Component Regression (PCR):**

  The team applies PCR to wine training data in order to estimate the regression coefficients in the model. In order to determine the best model, the team also calculates the test prediction errors when each component is applied.

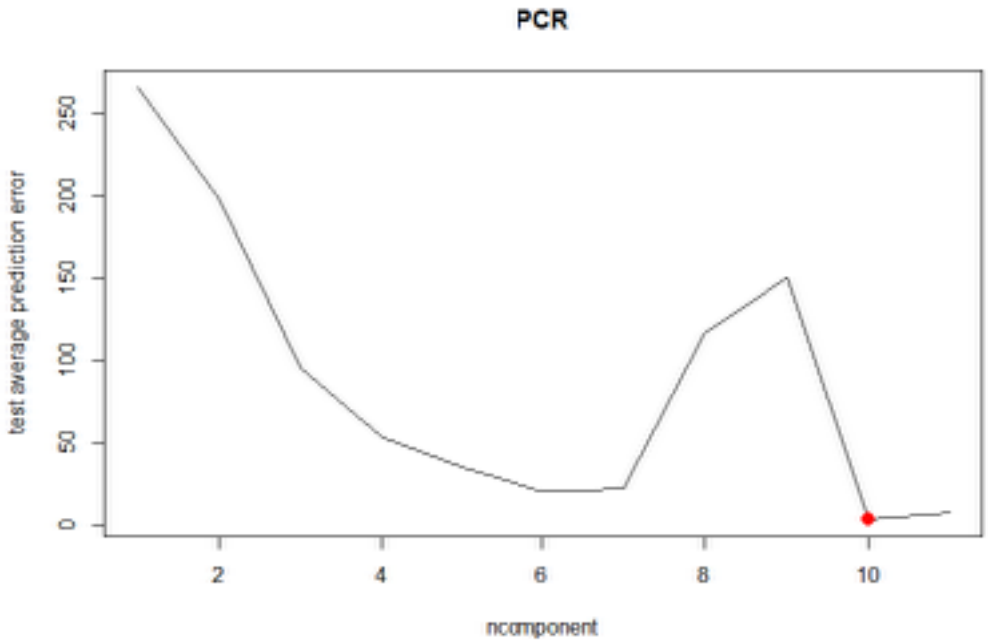

Figure 4: PCR cross-validation MSE

Then team finds out that the lowest cross-validation error occurs when M = 10 component is used, and **prediction error = 3.220977**; team can fit PCR on the full data set, using **M = 10.**

- **Partial Least Square Regression (PLS):**

  The team also applies PCL to wine training data in order to predict quality and evaluate the test set performance.

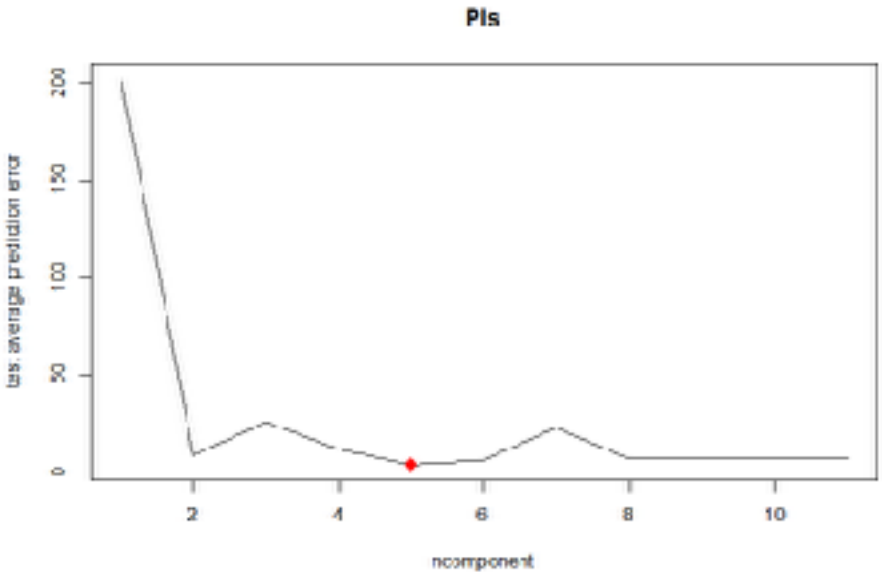

Figure 5: PLS cross-validation MSE

The team finds out the lowest cross-validation error occurs when M = 5 component is used and **prediction error = 3.796214**; team can fit PLSE on the entire data set, using **M = 5.**

5. <u>Support Vector Machines:</u>

The team randomly chooses 1000 samples from the original dataset to implement Support Vector Machine method since running the whole dataset keeps giving the warning message and is extremely slow. Because the data is highly unbalanced and there are only less than 1% data are in the quality which less than 3 and higher than 9, the team decides to remove the data in those ranges. The team then converts the quality of white wine to a categorical variable (3-5 for good, 6-9 for bad).

- **Linear Kernel**
  - The team first explores the nonnegative tuning parameter (cost of a violation to the margin) which allows the observations to be misclassified for linear SVM. The team first uses cross-validation to find best parameter by using linear kernel. From the method, observing from the best parameter is **cost = 10** and there are

479 support vectors from the summary of best model. Testing the data using the linear kernel with cost = 10 gives the following confusion matrix shown in Table 4. From the confusion matrix, the team calculated the test prediction accuracy which equals to **69.2%**.

Confusion Matrix		True Class		
		False	True	Total
Predicted Class	False	40	29	69
	True	48	133	181
	Total	88	162	250

Table 4: Confusion Matrix from Linear Kernel

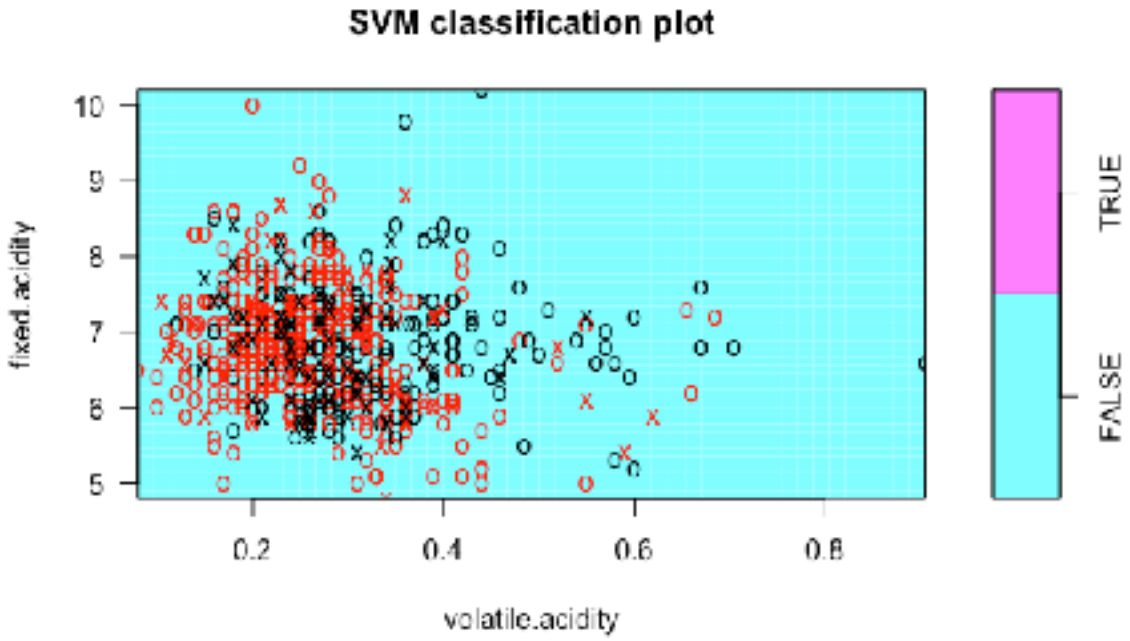

Figure 6: SVM (Linear) Classification plot

- **Radial Kernel**
  - The team then explores the non-negative tuning parameter C and the value of gamma for radial kernel SVM. The team uses cross-validation to find best

parameter. From the method, observing from the best parameter is **cost = 10, gamma = 1**. There are 634 support vectors from the summary of best model. Testing the data using the radial kernel with cost = 10 and gamma = 1 gives the following confusion matrix shown in Table 5. From the confusion matrix, the team calculated the test prediction accuracy which equals to **76%**.

Confusion Matrix		True Class		
		False	True	Total
Predicted Class	False	40	48	88
	True	12	150	162
	Total	52	198	250

Table 5: Confusion Matrix from Radial Kernel

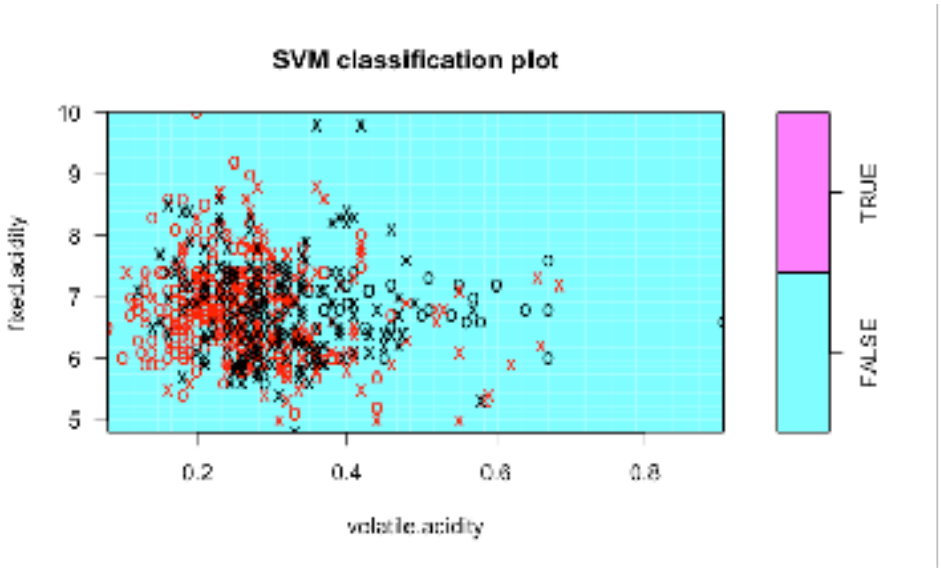

Figure 7: SVM (Radial) Classification plot

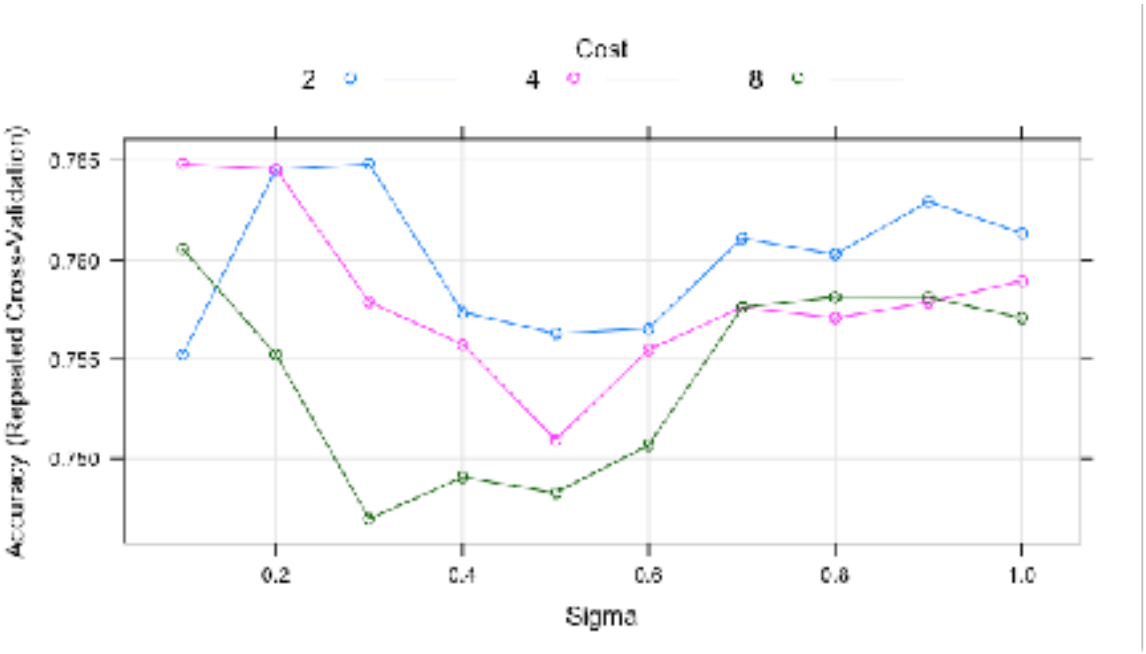

Figure 8: SVM (Radial) CV Accuracy

Radial kernel SVM has the highest accuracy. From figure 8, one can see that accuracy peaks at sigma = 0.3 with a cost value of 2.

- **Polynomial Kernel**
  - Lastly, the team explores the non-negative tuning parameter C and degree of freedom for polynomial kernel. The team uses cross-validation to find best parameter. From the method, observing from the best parameter is cost = 5, d=3 and there are 431 support vectors from the summary of best model. Testing the data using the radial kernel with **cost = 5 and d=3** gives the following confusion matrix shown in Table 6. From the confusion matrix, the team calculated the test prediction accuracy which equals to **72.4%**.

Confusion Matrix		True Class		
		False	True	Total
Predicted Class	False	34	15	49
	True	54	147	201
	Total	88	162	250

Table 6: Confusion Matrix from Polynomial Kernel

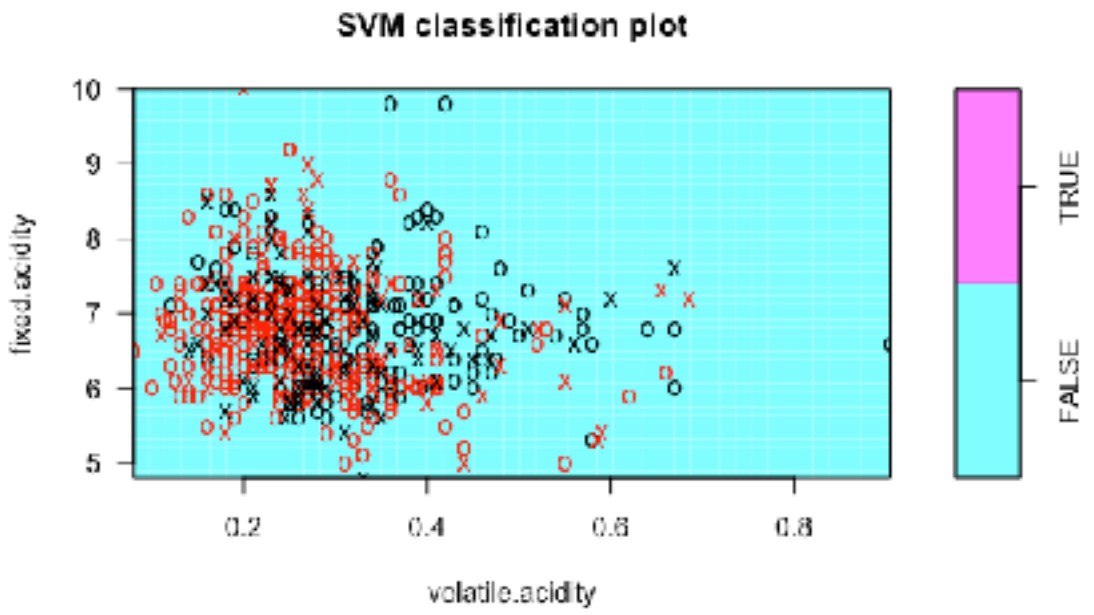

Figure 9: SVM (Polynomial) Classification plot

6. Neural Networks:

Neural Networks is used for classification. The team tunes parameters to best fit the cross-validation dataset through backward propagation. The team finds that **1 layer and 1 node** result in the lowest cross-validation error rate. The team generates the confusion matrix in Table 7. The accuracy rate is **80%**.

Figure 10: Neural Networks

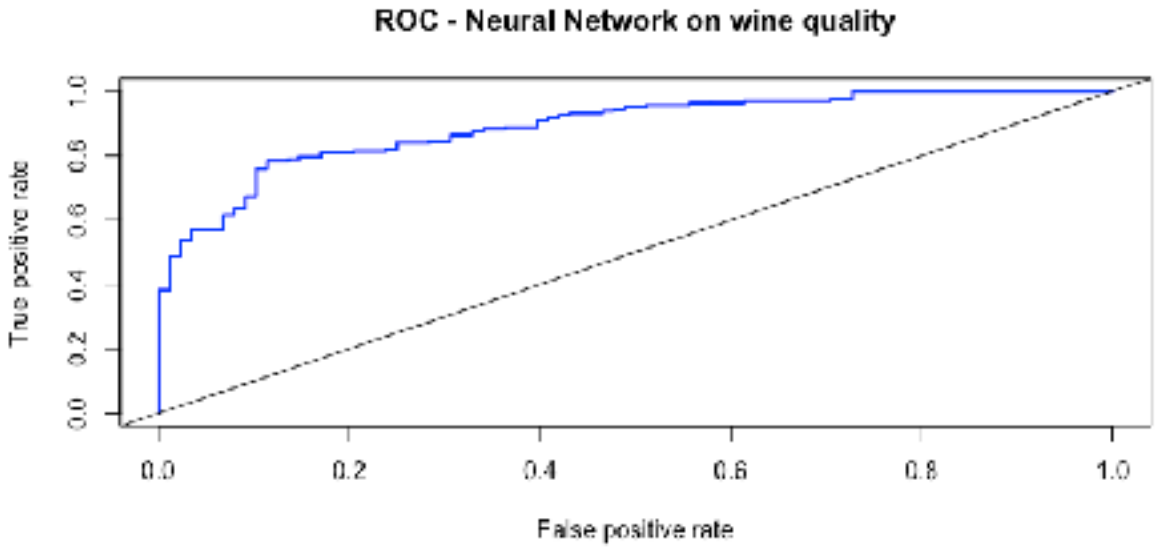

Figure 11: ROC Curve Neural Networks

Confusion Matrix		True Class		
		False	True	Total
Predicted Class	False	61	27	88
	True	23	139	162
	Total	84	166	250

Table 7: Confusion Matrix from Neural Network

7. Random Forest

For the random forest, the team uses ntree and mtry to specify the total number of trees to build (default = 500), and the number of predictors to randomly sample at each split respectively. Mtry is the number of variables randomly sampled as candidates at each split. 500 trees were built, and the model randomly sampled 3 predictors at each split. Accuracy peaks at **mtry = 1**. From confusion matrix, the team generates the accuracy rate which equals to **82.4%**.

Confusion Matrix		True Class		
		False	True	Total
Predicted Class	False	59	10	69
	True	34	147	181
	Total	93	157	250

Table 8: Confusion Matrix from Random Forest

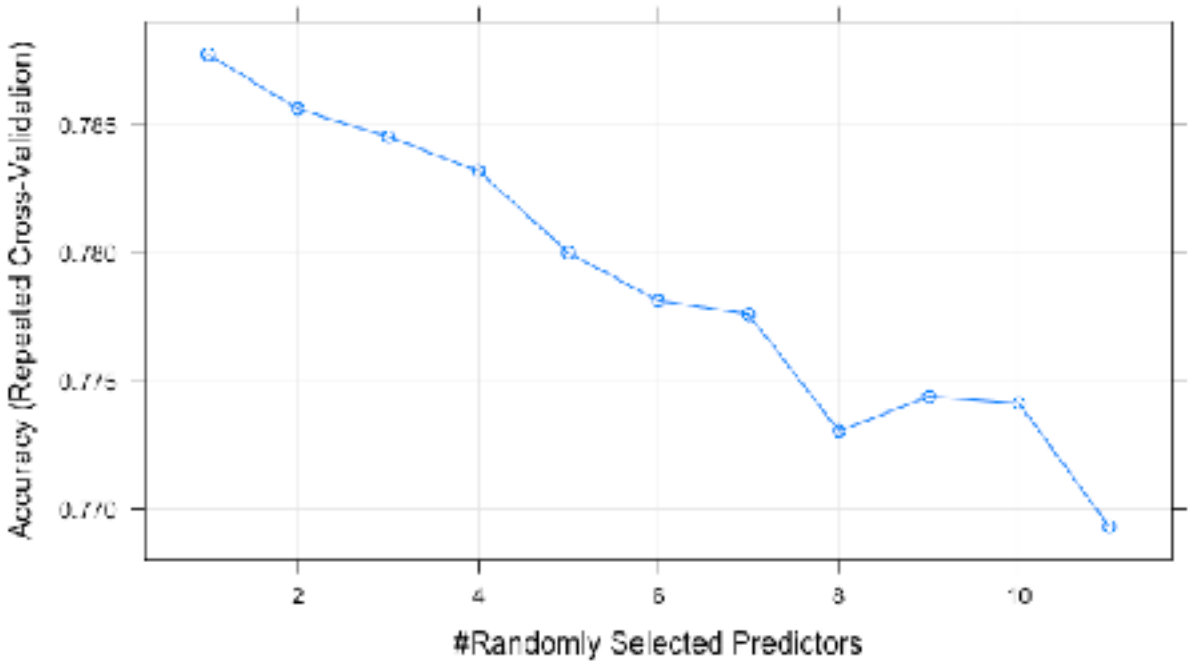

Figure 12: Random Forest cross-validation MSE

Discussion:

1. **How human quality of tasting is related to the chemical properties of wine? Discuss the result from best subset, stepwise selection, shrinkage and dimension reduction methods.**

   The plot of RSS, adjusted R^2, Cp, and BIC for all of the models using best subset methods is shown in Figure 1. One can see that different criteria can indeed lead to different "best" models.The team chooses the final model based on the test error using the validation set and cross-validation methods, the team is going to select the model for which the resulting estimated test error is smallest. For the wine data set, applying the validation rule leads the team to the selection of eight-variable model. The testing error associated with this model is **0.5973154**. Once the model is selected, the team applies the model on the entire wine data set, the final model is given below:
   Quality = (Intercept) +B1 fixed.acidity + B2 volatile.acidity +B3 residual.sugar + B4 free.sulfur.dioxide + B5 density + B6 ph + B7 sulphates + B8 alcohol, where,

Intercept	B1	B2	B3	B4
1.541062e+02	6.810394e-02	-1.888140e+00	8.28472e-02	3.349015e-03
B5	B6	B7	B8	
-1.54291e+02	6.942135e-01	6.285081e-01	1.93168e-01	

   The best subset, forward stepwise, backward and hybrid stepwise selection approaches generally give similar but not identical models. However, for the wine data set, the team finds that these methods yield the same best models. The test prediction error associated with forward, backward and hybrid stepwise selection is **0.5973154.**
   Based on best subset and stepwise selection, the quality of white wine is related to **eight chemical properties, which are fixed.acidity, volatile.acidity, residual.sugar, free.sulfur.dioxide, density, ph, sulphates and alcohol.** Based on the coefficients from the table above, density has the greatest influence on wine quality assessment. The **density** depends on the percent alcohol and sugar content and it will directly affect the

taste of the wine. This could somehow explain why when there is one unit change in density, the response variable will be alter by -154.29.

Both ridge and lasso regression suggest a model with **all the eleven attributes.** Lasso regression fails to shrink the effect of any attribute to zero for the wine data set. Therefore the final model from shrinkage contains all the attributes in the data set. Ridge and lasso regression generate different regression coefficients on the entire wine data set, but both final equations show that density has the greatest impact on the wine quality assessment compared to other chemical properties.

The team further uses principal component regression (PCR) and partial least squared (PLS). The wine data set contains both numerical and categorical variables. Applying PCA directly on categorical variables may be problematic and biased, but the team still decides to use PCA, not expecting to see PCA could predict wine quality better than other regression methods do. Instead, the team is interested in how PCA is going to reduce the dimension for the wine data set. In resulting of choosing the principal component that generate the lowest prediction error, the team chooses M=10 for PCR and M=5 for plse.

The team looks at the component loadings for PCR when M =10. Component loadings will tell how much of the variation in a variable will be explained by the component. The highest component loading for PCR when component = 10 is free.sulfur. Dioxide and for PLS when component 5 is fixed.acidity. Large loadings indicate that a variable has a strong effect on the principal component.

2. Which regression methods give more desirable results for wine data set? Make comparison between best models.

Compare best model between regression types:

	Test Prediction Error	Variables used
Best Subset	**0.5973154**	fixed.acidity + volatile.acidity + residual.sugar + free.sulfur.dioxide + density + ph + sulphates + alcohol

Stepwise	**0.5973154**	fixed.acidity + volatile.acidity +B3 residual.sugar + free.sulfur.dioxide + density + ph + sulphates + alcohol
Ridge Regression	**0.5969285**	**All eleven variables**
Lasso Regression	0.5973147	**All eleven variables**
PCR	3.220977 (M=10)	**All eleven variables**
PLSE	3.796214 (M=5)	**All eleven variables**

The best model ridge regression yields the lowest test average prediction error among all the regression types. Since ridge regression doesn't select variables nor remove attributes from the full model, it does not surprise us that ridge regression yields the lowest test average prediction error. The same reason applies to lasso regression. None of the estimates of chemical properties are shrunk to zero by lasso regression; the lasso regression results in a full model as well. Best subset and stepwise selection also yield a relatively low prediction error and the best model suggested contain 8 variables.

Principal Component Analysis does not work really well in predicting the wine quality. The response variable is discrete integer, usually regular PCA on the raw data is not recommended because the distance between categories may be problematic and may not be reasonable. However , the team is still interested in how PCA does its job in dimension reduction. The team first applies PCR on the f data set and calculates the prediction error associated with each number of principal component. From figure.4 one might see that the lowest test prediction error occur when M = 10 component is used and prediction error is 3.220977. As mentioned before, the prediction error using PCR is much bigger than best subset and stepwise selection.

The team also applies PLS on the training data set and calculates the prediction error. From figure 5, one might suggest that the lowest test prediction error occur when M=5. The prediction error associated with M = 5 is 3.796214.Based on the lowest prediction

error, the team chooses component 10 for PCR and component 3 for PLS and calculates the corresponding regression coefficients.

In conclusion, even though the best model from ridge and lasso regression yield the lowest test average prediction error among all the regression methods, this is because these two methods include all the attributes in the model. If one wants to reduce the number of explaining variables in the model, both best subset and stepwise selection yield prediction error close to ridge and lasso. The best model from best subset and stepwise contain eight variables: fixed.acidity, volatile.acidity, residual.sugar, free.sulfur.dioxide, density, ph, sulfates and alcohol.

3. **Can we classify good or bad quality of white wine using the 11 properties of white wine? Applying three models, random forest, Neural Networks, and Support Vector Machines, on the dataset to illustrate which performance of model type is better.**

The team conducts the classification analysis using Linear SVM, Radial SVM, Polynomial SVM, Neural Networks and Random Forest:

	Accuracy	Specificity	Sensitivity
SVM (linear)	69.2%	45.45%	82.10%
SVM (radial)	76.0%	76.92%	75.76%
SVM (polynomial)	72.4%	38.64%	90.74%
Neural Networks	80.0%	72.62%	83.73%
Random Forest	82.4%	63.44%	93.63%

Accuracy is the proportion of true results, either true positive or true negative, in a population. It measures the degree of veracity of the classification, and is calculated as (TN + TP)/(TN+TP+FN+FP) = (Number of correct assessments)/Number of all assessments). Sensitivity is the proportion of true positives that are correctly identified by

the classification. It shows how good the model is at detecting high quality of white wine. Specificity is the proportion of the true negatives correctly identified by the classification. It suggests how good the model is at identifying low quality of white wine.

From the summary table , one can see that s. Random Forest is overall the best model in classifying high and low quality of white wine. However, when specifically detecting low quality of white wine, Neural Networks yields better result. Therefore the team recommends using Random Forest for general classification question and Neural Networks for identifying low quality of white wine.

Conclusion:

The results of the project do provide some important discoveries for the wine industry. Yet, the evaluations (qualities) are based in the experience and knowledge of the experts, which are prone to subjective factors. The proposed data-driven approach the team uses is based on objective tests. Although It's hard to accurately predict subjective results through objective tests, the methods using in the project have offered some reasonably valid approaches.

Citation

P. Cortez, A. Cerdeira, F. Almeida, T. Matos and J. Reis.

Modeling wine preferences by data mining from physicochemical properties. In Decision Support Systems, Elsevier, 47(4):547-553, 2009.

Portilla, Jose. "KDnuggets." *KDnuggets Analytics Big Data Data Mining and Data Science*. KDnuggets, 2016. Web. 05 Dec. 2016.

"Neural Network Data Mining Explained - Butler Analytics." *Butler Analytics*. Butler, 15 June 2015. Web. 05 Dec. 2016.

Benyamin, Dan. "A Gentle Introduction to Random Forests, Ensembles, and Performance Metrics in a Commercial System." *Blog & Press*. N.p., 02 Aug. 2014. Web. 05 Dec. 2016.