CS 545 ASSIGNMENT 2

Paresh Bhambhani

October 2, 2015

Contents

1 PART 1

1.1 RMSE and MAD

The Ridge regression implementation can be seen in the attached python file (Due to the length of the code, it has not been added in the appendix as it was going over the 2 page limit). RMSE and MAD are computed for ridge regression applied to the standardized Red and White wine data set. Figure 1 shows the plot for RMSE and MAD for the Red wine data set and Figure 2 for the White wine data set. After $\lambda = 100$ both RMSE and MAD show a jump in the error rate. Based on the graphs we can see that $\lambda = 100$ is a good choice.

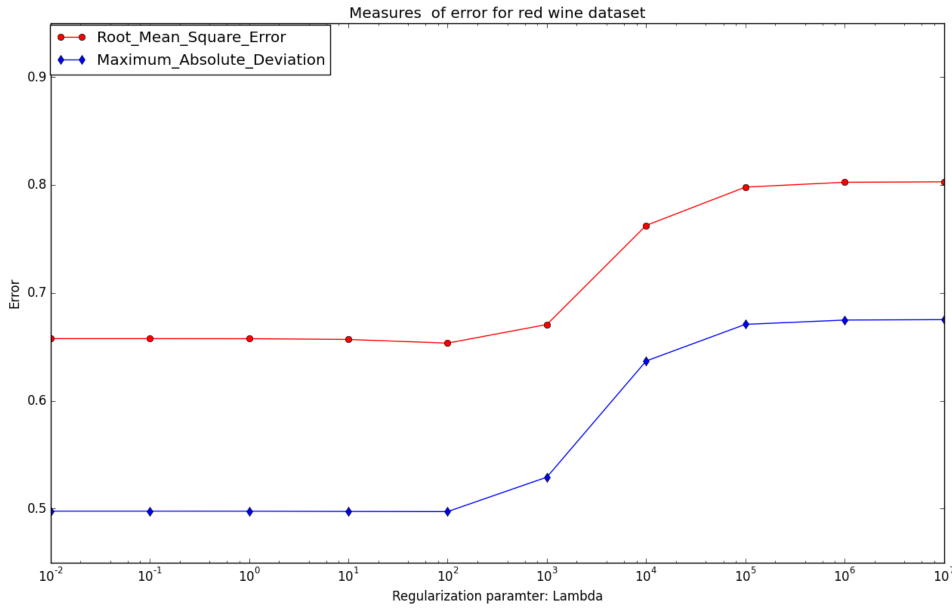

Figure 1: Measure of Error for Red Wine dataset

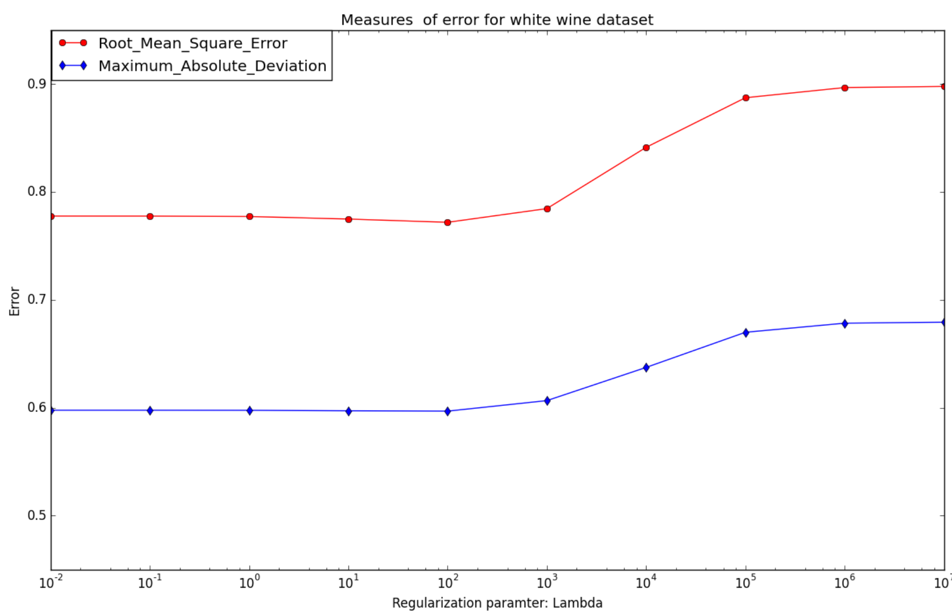

Figure 2: Measure of Error for White Wine dataset

1.1.1 Why MAD over RMSE?

To Answer this question let us consider a hypothetical case. Let the error values between the actual value and predicted value be denoted as $\mathbf{e}_1, \mathbf{e}_2, \mathbf{e}_3, \mathbf{e}_4$. where $\mathbf{e}_i$ is

$$\mathbf{e}_i = y_i - h(x_i)$$

Now we see the effect of increasing the variance of the magnitudes of these errors on MAD and RMSE as illustrated in Table 1

Table 1: Hypothetical case with 4 errors and their RMSE and MAD

Variable	Case 1	Case 2	Case3
$\mathbf{e}_1$	2	1	0
$\mathbf{e}_2$	2	1	0
$\mathbf{e}_3$	2	1	0
$\mathbf{e}_4$	2	5	8
MAD	2	2	2
RMSE	2	2.6	4

As is clear from the table, as the magnitude of any of the error gets large, the RMSE blows over but the MAD still retains its values. This comes into play in the case when outliers are present in the data. Since outliers have large magnitude of error, they affect the RMSE by increasing it. Although it will not affect greatly if we have large number of data points and only few outliers since we normalize by **N** (total number of data points), but the effect becomes apparent if we have more outliers. As seen above this however does not affect the MAD.

1.2 The REC Curve

The REC algorithm was implemented as explained in [1]. Figure 3 shows the REC curve for red and white wine using absolute deviation as done in [2]. As mentioned above, λ is chosen as 100.

As seen from the graphs, the curve for the red wine is generally higher than the curve for the white wine thus indicating that the Ridge regression works better for the Red wine data set compared to the white wine data set.
REC curves can give a visual idea of how good (or bad) the regression model is doing. Unlike RMSE or MAD, REC curves for various models can indicate which model is working the best on the given data.The curve area provides a measure of the performance of the regression model which cannot be gauged from RMSE and MAD graphs.

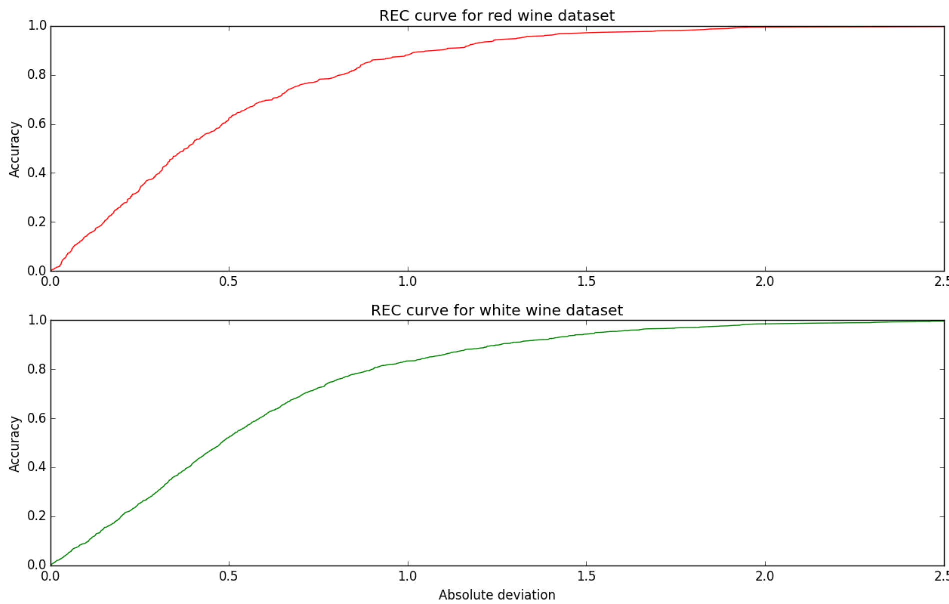

Figure 3: REC Curves for Red and White wine data sets

1.3 Comparing our results with the Paper

To compare our results with that of the paper, refer Table 2. We compare our ridge regression across other models discussed in [2].

Table 2: Comparison of Algorithms

Wine	Parameter	RR	MR	NN	SVM
Red Wine	MAD	0.49	0.50	0.41	0.46
	$Accuracy_{\tau=0.25}$	31.59	31.2	31.1	43.2
	$Accuracy_{\tau=0.50}$	60.24	59.1	59.1	62.4
	$Accuracy_{\tau=1.0}$	87.6	88.6	88.8	89.0
White Wine	MAD	0.59	0.59	0.48	0.45
	$Accuracy_{\tau=0.25}$	25.26	25.6	26.5	50.3
	$Accuracy_{\tau=0.50}$	51.84	51.7	52.6	64.6
	$Accuracy_{\tau=1.0}$	83.3	84.3	84.7	86.8

The table compares our Ridge regression (RR) against Multiple Regression (MR), Neural Network (NN), and Support Vector Machine (SVM). It can be seen from the table that Ridge regression gives similar results as the Multiple regression approach in the paper. However SVM gives the best results.

2 Part 2

2.1 Relationship between the magnitude of weight vector components and their relevance to the classification task

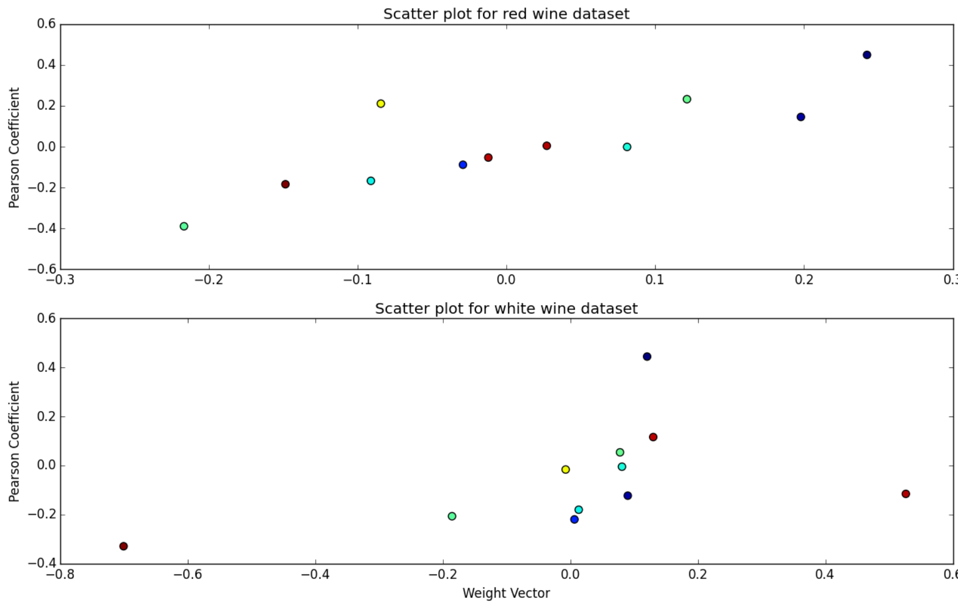

Figure 4: Scatter Plots of subsets of Red and White wine Training datasets

Figure 4 show scatter plots of the feature weight vector against the Pearson Correlation Coefficient of features on 50% training dataset size. Ideally this plot should be along the diagonal, increasing as we move towards right which is generally the case here. Higher correlation of the points down right shows that the weight vector components correspond to higher correlation. Points that are off from the diagonal indicate that PCC does not correspond with the weight. This may be an indication of the feature being incorrectly weighed.
The Table 3 compares the ranks of features obtained from Ridge regression with those obtained in [2]. To get the ranks for ridge regression, I marked the values of vector weight components with first four highest Pearson coefficients and then mapped those vector components to the corresponding features.

Table 3: Comparison of Feature Ranks

Wine	RR	SVM
Red Wine	Alcohol	sulphates
	Citric Acid	pH
	Sulphates	total sulphur dioxide
	Fixed Acidity	alcohol
White Wine	Alcohol	sulphates
	pH	alcohol
	Sulphates	residual sugar
	free sulphur dioxide	citric acid

2.2 Incremental Feature Removal

In this section we incrementally remove the feature with the lowest absolute value of the weight vector component and plot the RMSE and MAD errors against the number of features. Figure 5 shows the described graphs.

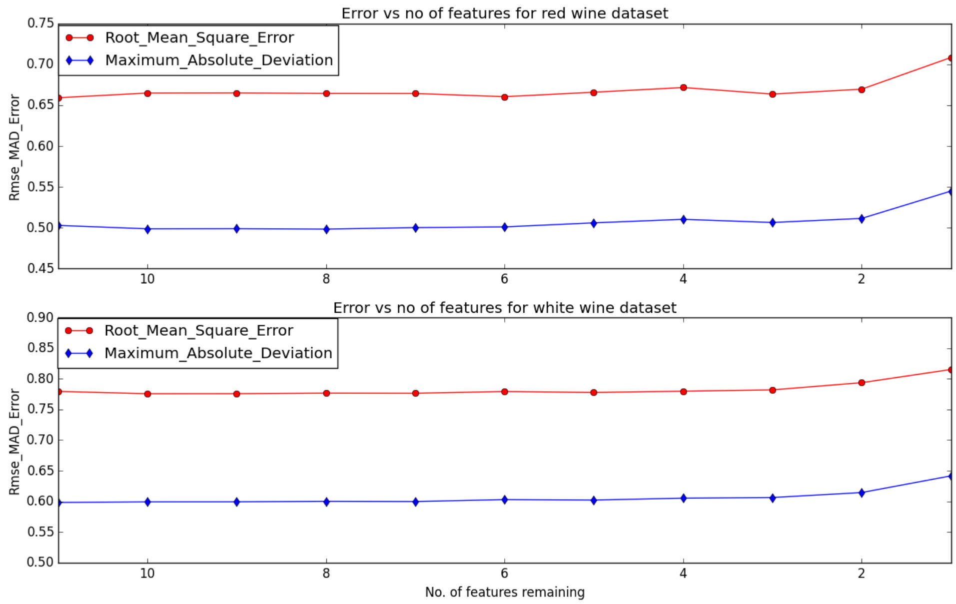

Figure 5: Effect of Feature Removal on error

If we remove a feature that is not very important, it will not affect the overall accuracy much however if we remove an important feature, it will result in an increase in the error rate. As can be seen from the figure, removal of the first few features has little effect on the overall accuracy. We can see the curve hardly rising for subsequent feature removals. For both Red and White wines, removal of the 6th feature results in a jump in the error rate which then increases with every removal.

References

[1] Jinbo Bi BIJ, RPI Edu, and Kristin P Bennett BENNEK. Regression error characteristic curves.

[2] Paulo Cortez, António Cerdeira, Fernando Almeida, Telmo Matos, and José Reis. Modeling wine preferences by data mining from physicochemical properties. *Decision Support Systems*, 47(4):547–553, 2009.