Yianni Anastopoulos
Ryan Hausen
Mini Project

The document is broken in to sections delineated by bold fonts with horizontal lines:
- Datasets
- Algorithms
- Initial Submission
- updates...

Datasets

Regression Datasets:

UCI ML – Abalone

Sex / nominal / -- / M, F, and I (infant)
Length / continuous / mm / Longest shell measurement
Diameter / continuous / mm / perpendicular to length
Height / continuous / mm / with meat in shell
Whole weight / continuous / grams / whole abalone
Shucked weight / continuous / grams / weight of meat
Viscera weight / continuous / grams / gut weight (after bleeding)
Shell weight / continuous / grams / after being dried
Rings / integer / -- / +1.5 gives the age in years

UCI ML – Forest Fires

For more information, read [Cortez and Morais, 2007].
1. X - x-axis spatial coordinate within the Montesinho park map: 1 to 9
2. Y - y-axis spatial coordinate within the Montesinho park map: 2 to 9
3. month - month of the year: 'jan' to 'dec'
4. day - day of the week: 'mon' to 'sun'
5. FFMC - FFMC index from the FWI system: 18.7 to 96.20
6. DMC - DMC index from the FWI system: 1.1 to 291.3
7. DC - DC index from the FWI system: 7.9 to 860.6
8. ISI - ISI index from the FWI system: 0.0 to 56.10
9. temp - temperature in Celsius degrees: 2.2 to 33.30
10. RH - relative humidity in %: 15.0 to 100
11. wind - wind speed in km/h: 0.40 to 9.40
12. rain - outside rain in mm/m2 : 0.0 to 6.4
13. area - the burned area of the forest (in ha): 0.00 to 1090.84

(this output variable is very skewed towards 0.0, thus it may make sense to model with the logarithm transform).

UCI ML – Wine Quality

For more information, read [Cortez et al., 2009].
Input variables (based on physicochemical tests):
1 - fixed acidity
2 - volatile acidity
3 - citric acid
4 - residual sugar
5 - chlorides
6 - free sulfur dioxide
7 - total sulfur dioxide
8 - density
9 - pH
10 - sulphates
11 - alcohol
Output variable (based on sensory data):
12 - quality (score between 0 and 10)

Classification Datasets

UCI ML – Iris

1. sepal length in cm
2. sepal width in cm
3. petal length in cm
4. petal width in cm
5. class:
-- Iris Setosa
-- Iris Versicolour
-- Iris Virginica

UCI ML – Wine Quality
Input variables (based on physicochemical tests):
1 - fixed acidity
2 - volatile acidity
3 - citric acid
4 - residual sugar
5 - chlorides
6 - free sulfur dioxide
7 - total sulfur dioxide
8 - density
9 - pH

10 - sulphates
11 - alcohol
Output variable (based on sensory data):
12 - quality (score between 0 and 10)

UCI ML – Abalone

Sex / nominal / -- / M, F, and I (infant)
Length / continuous / mm / Longest shell measurement
Diameter / continuous / mm / perpendicular to length
Height / continuous / mm / with meat in shell
Whole weight / continuous / grams / whole abalone
Shucked weight / continuous / grams / weight of meat
Viscera weight / continuous / grams / gut weight (after bleeding)
Shell weight / continuous / grams / after being dried
Rings / integer / -- / +1.5 gives the age in year

Algorithms

Regression:

- Linear Regression
- Decision Tree Regression
- KNN
- SVM Regressor
- Lasso (11/8/2018)
- Ridge (11/8/2018)

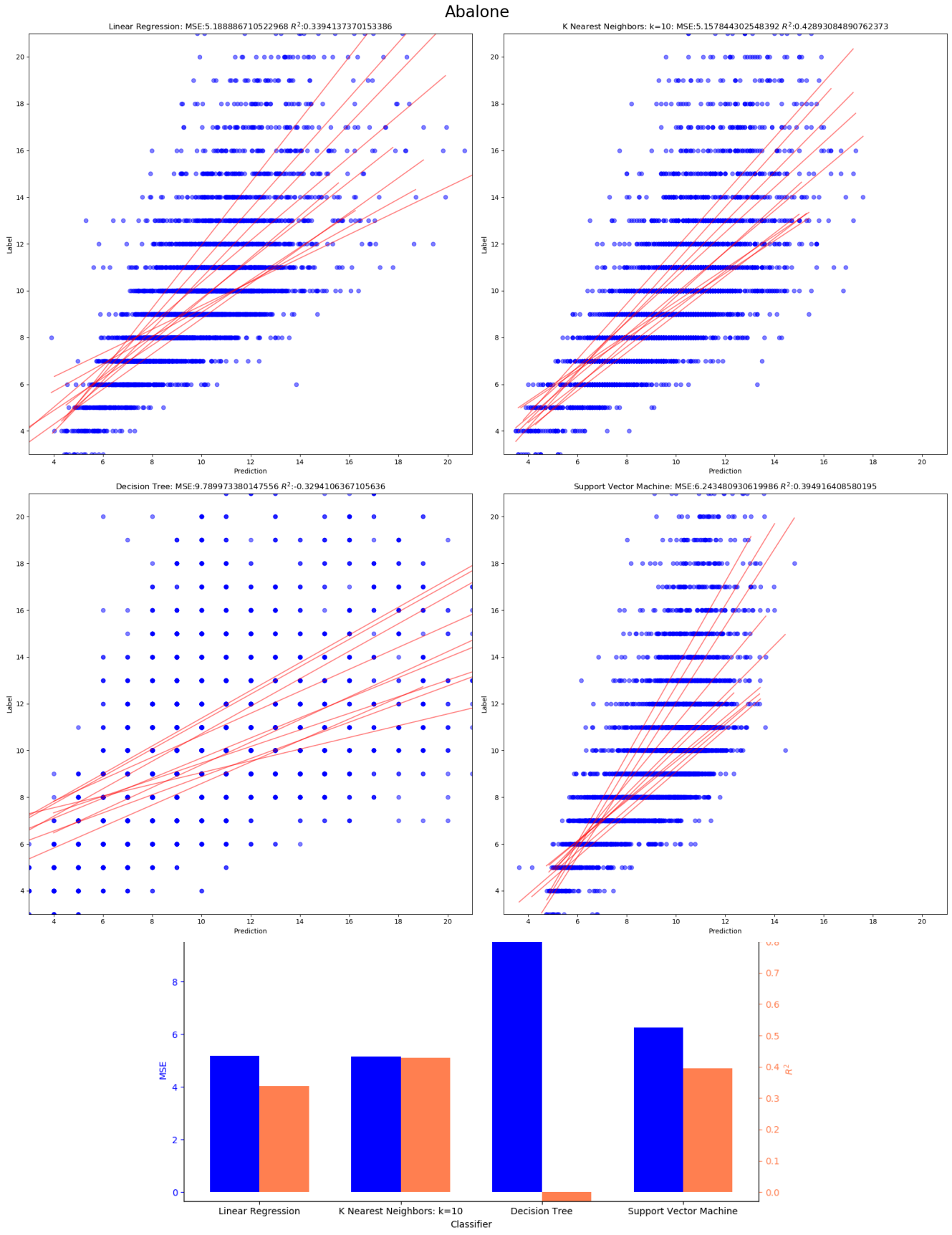
Classification:

- Logistic Regression
- Decision Tree Classifier
- KNN
- LDA & QDA
- SVM

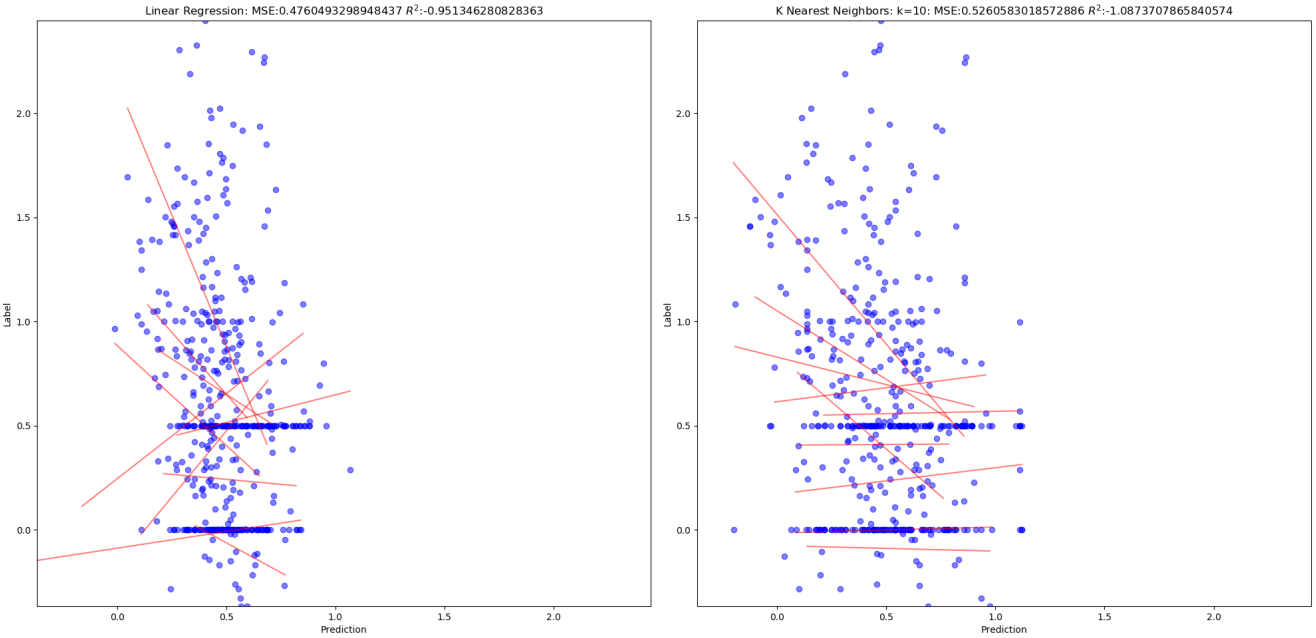
Initial Submission

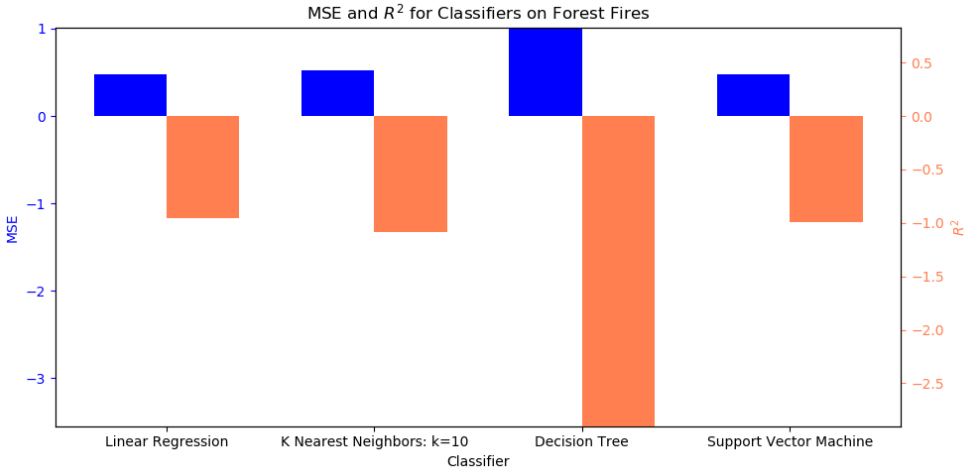
Regression Results

All crossvalidation predictions, k=10, with a fitted line for each set for k

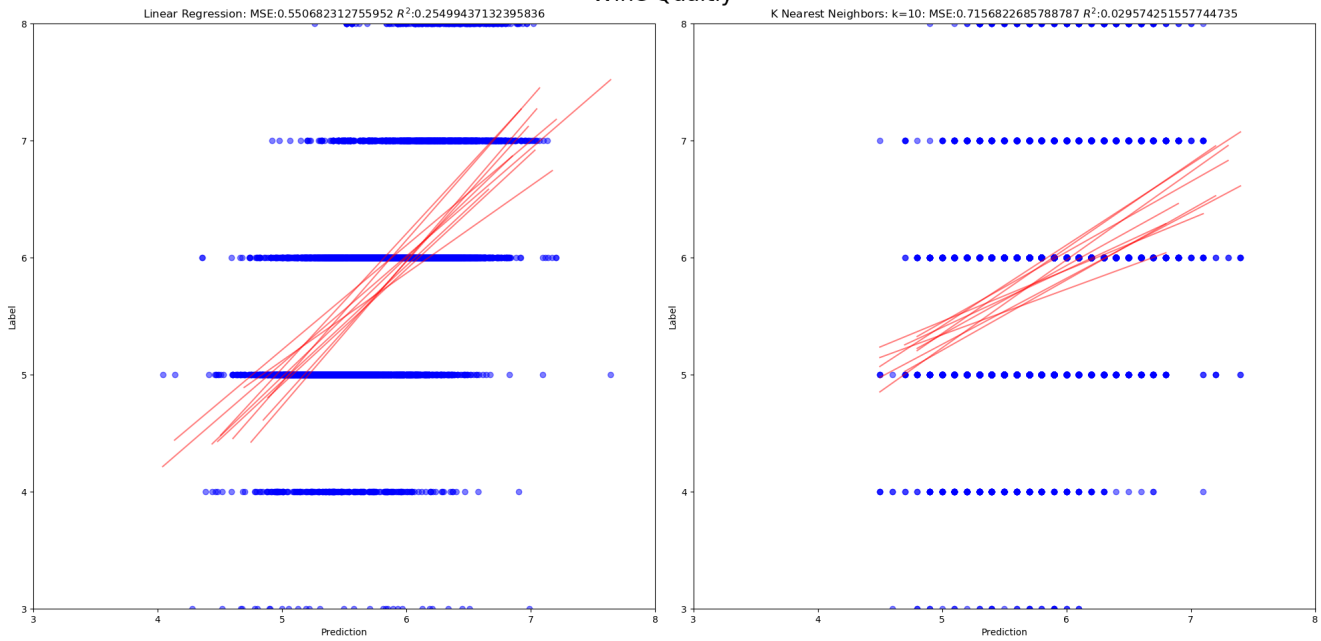


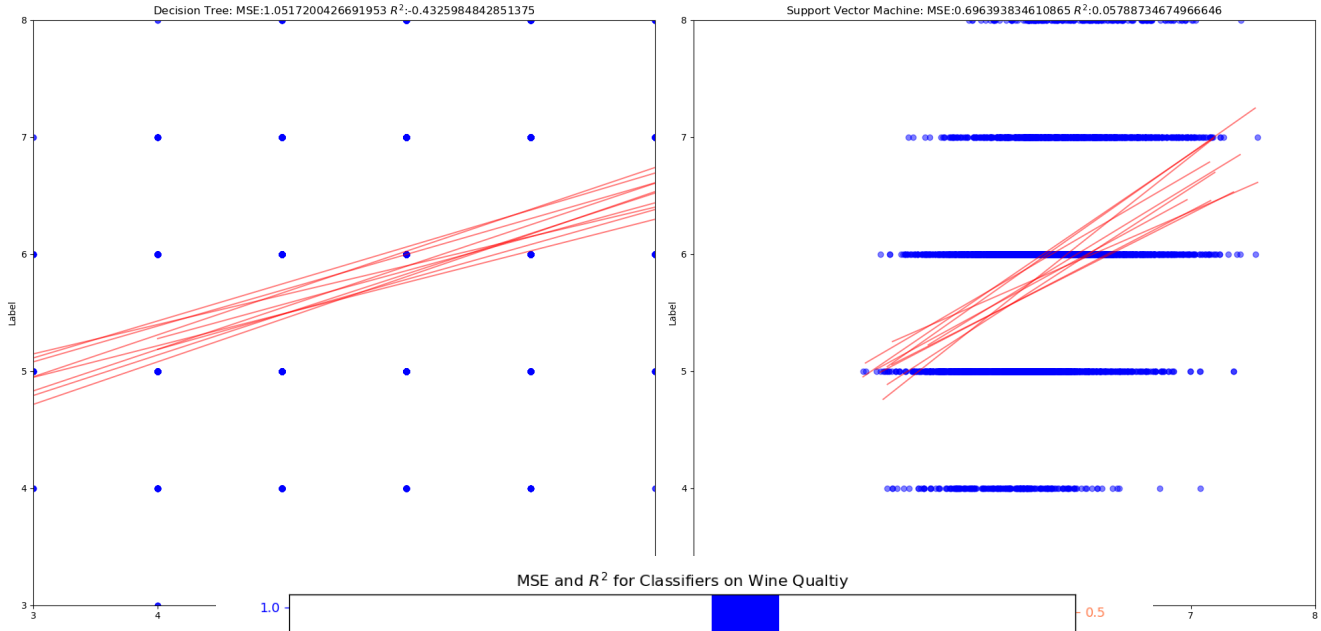
Forest Fires

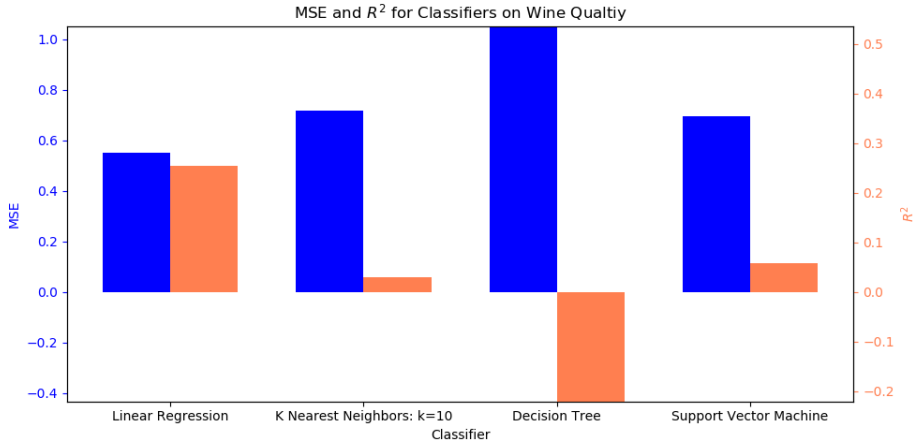




Wine Qualtiy

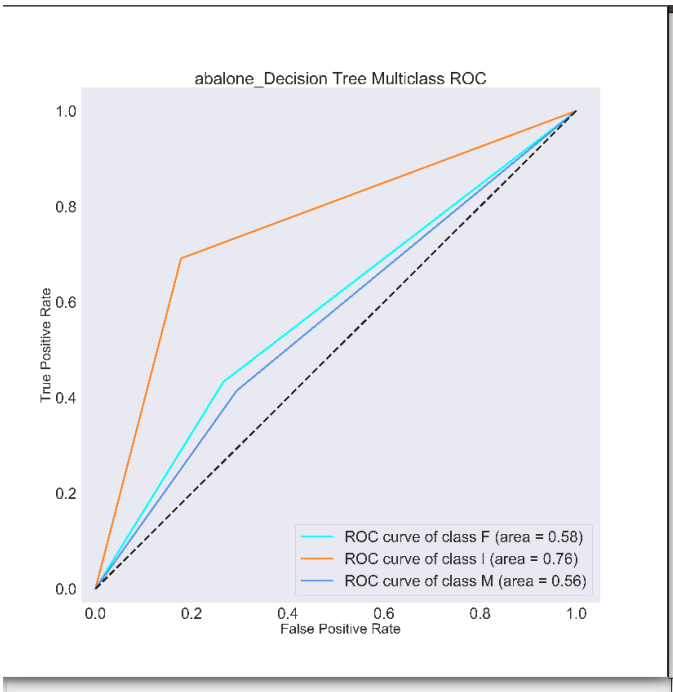


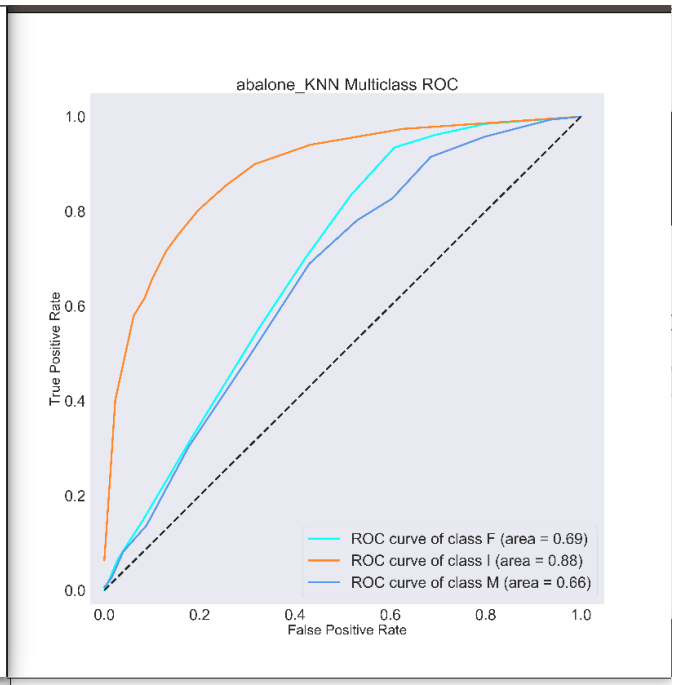


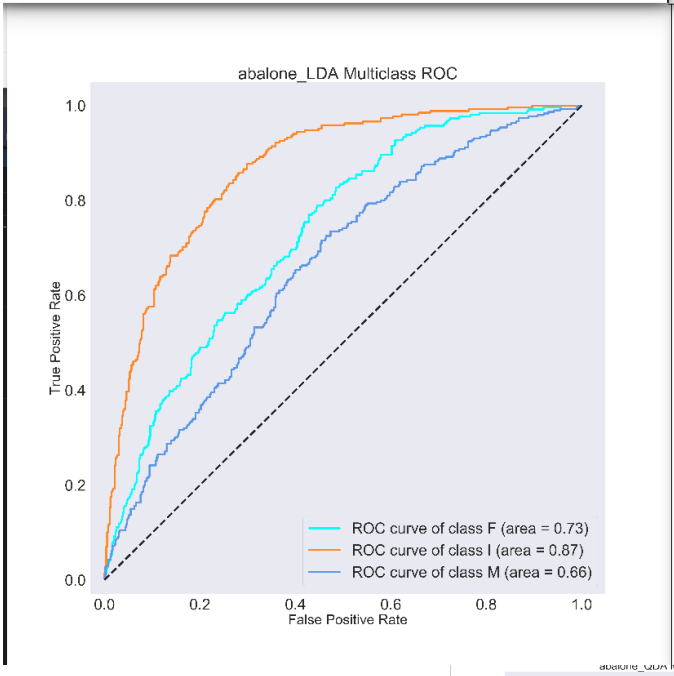


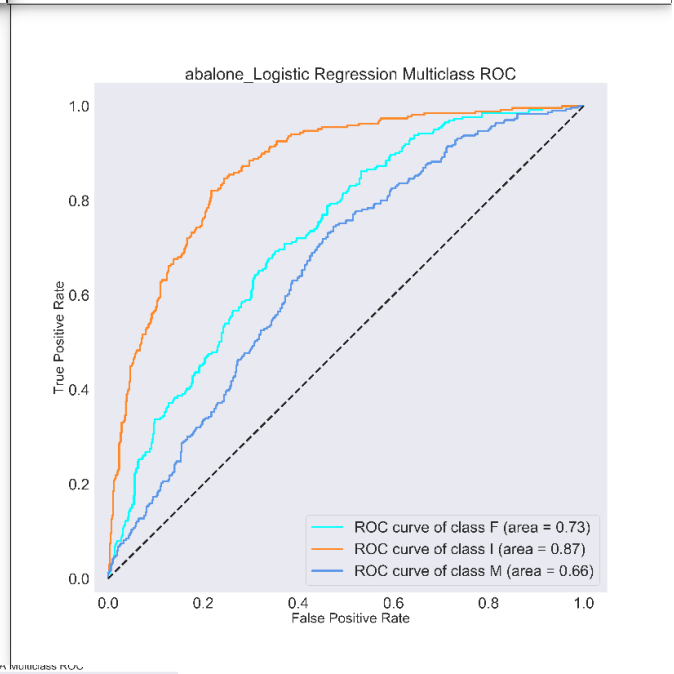
Classification Results

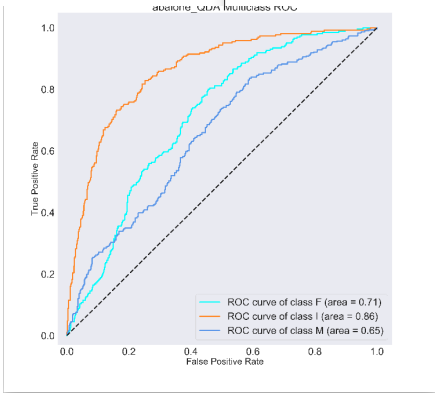
Abalone



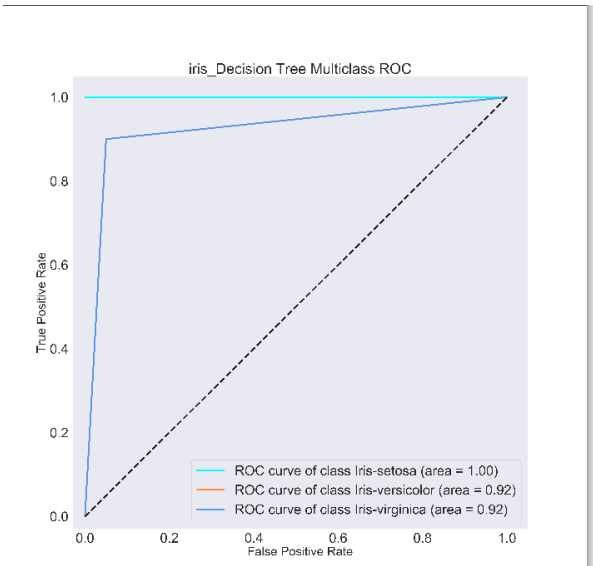


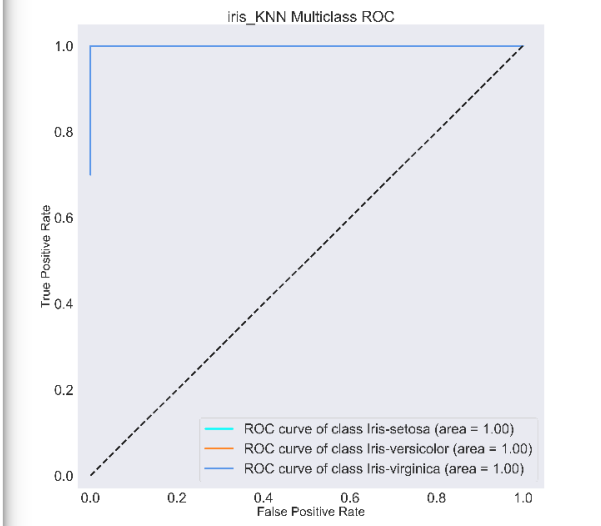


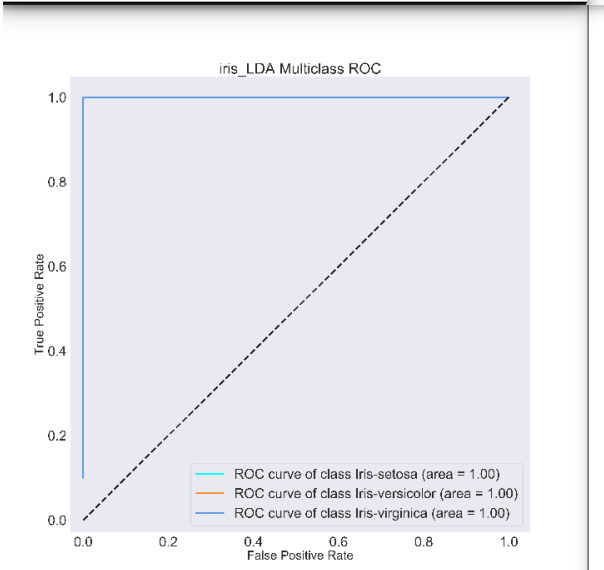


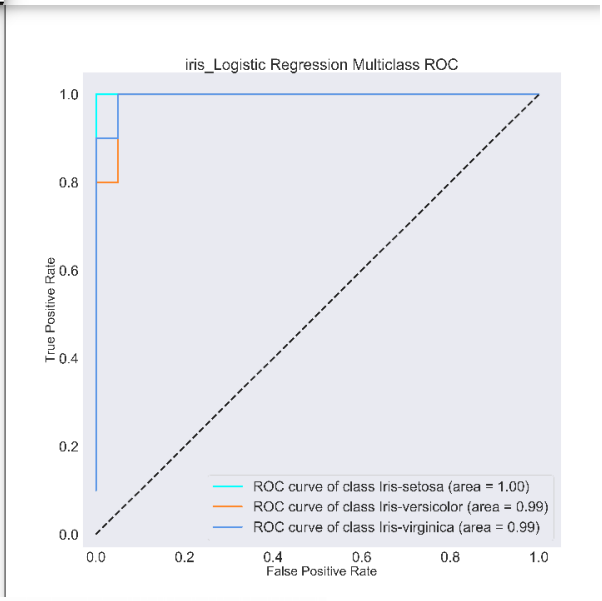


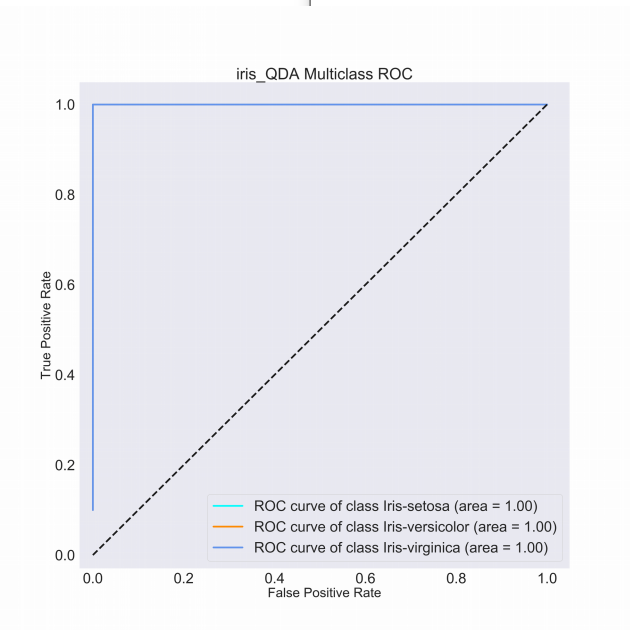
Iris


iris_Decision Tree Multiclass ROC
True Positive Rate
False Positive Rate
ROC curve of class Iris-setosa (area = 1.00)
ROC curve of class Iris-versicolor (area = 0.92)
ROC curve of class Iris-virginica (area = 0.92)


iris_KNN Multiclass ROC
True Positive Rate
False Positive Rate
ROC curve of class Iris-setosa (area = 1.00)
ROC curve of class Iris-versicolor (area = 1.00)
ROC curve of class Iris-virginica (area = 1.00)

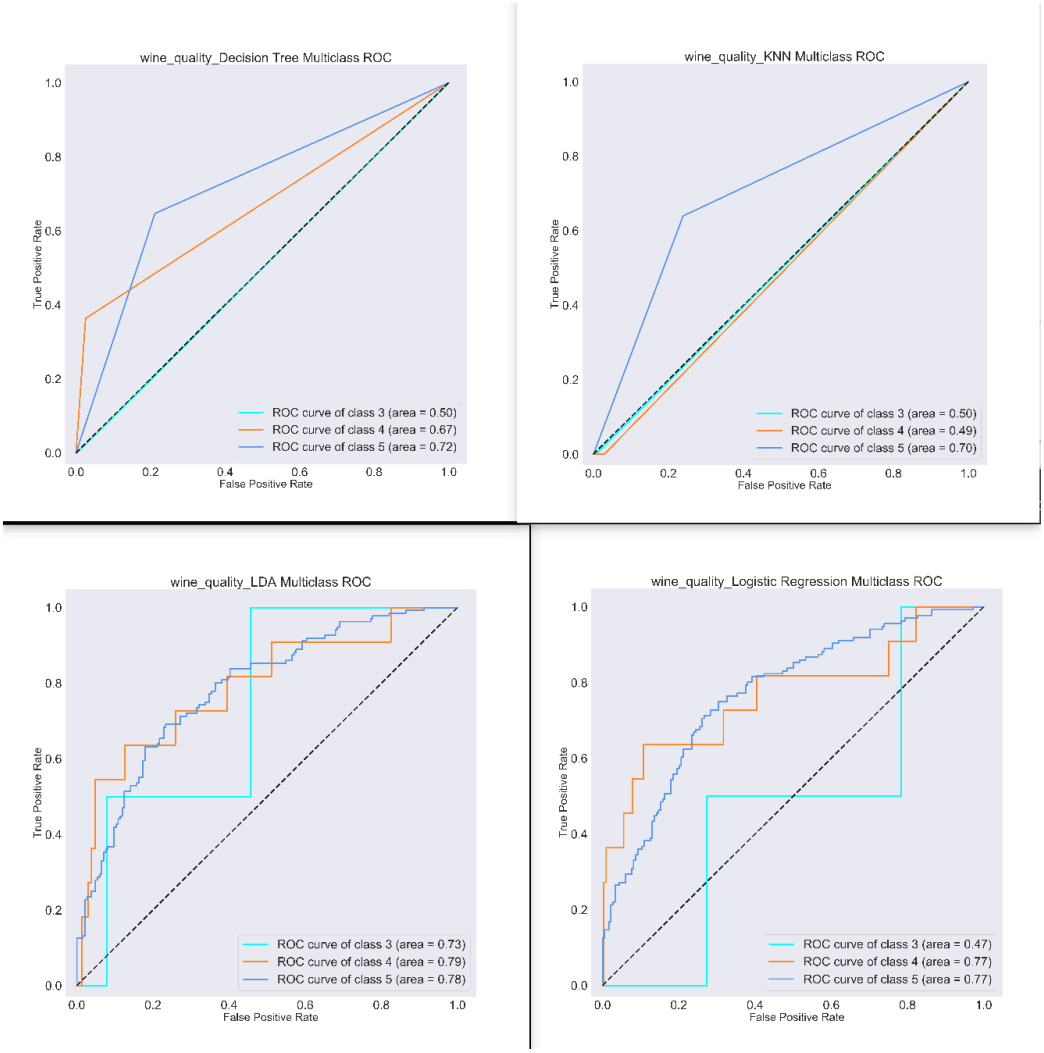

iris_LDA Multiclass ROC
True Positive Rate
False Positive Rate
ROC curve of class Iris-setosa (area = 1.00)
ROC curve of class Iris-versicolor (area = 1.00)
ROC curve of class Iris-virginica (area = 1.00)

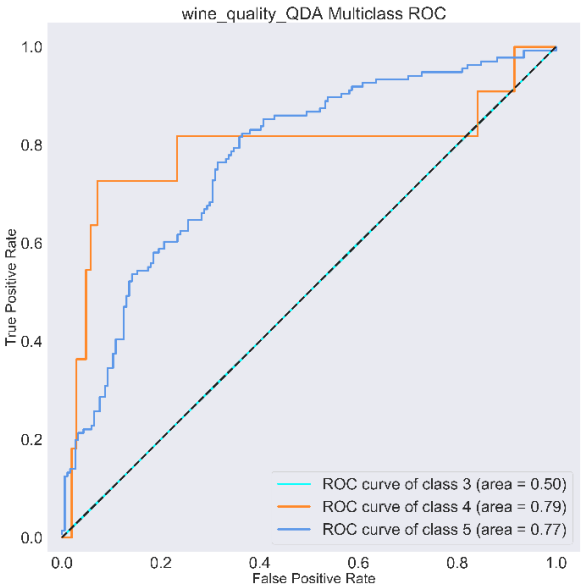

iris_Logistic Regression Multiclass ROC
True Positive Rate
False Positive Rate
ROC curve of class Iris-setosa (area = 1.00)
ROC curve of class Iris-versicolor (area = 0.99)
ROC curve of class Iris-virginica (area = 0.99)


iris_QDA Multiclass ROC
True Positive Rate
False Positive Rate
ROC curve of class Iris-setosa (area = 1.00)
ROC curve of class Iris-versicolor (area = 1.00)
ROC curve of class Iris-virginica (area = 1.00)

.any()

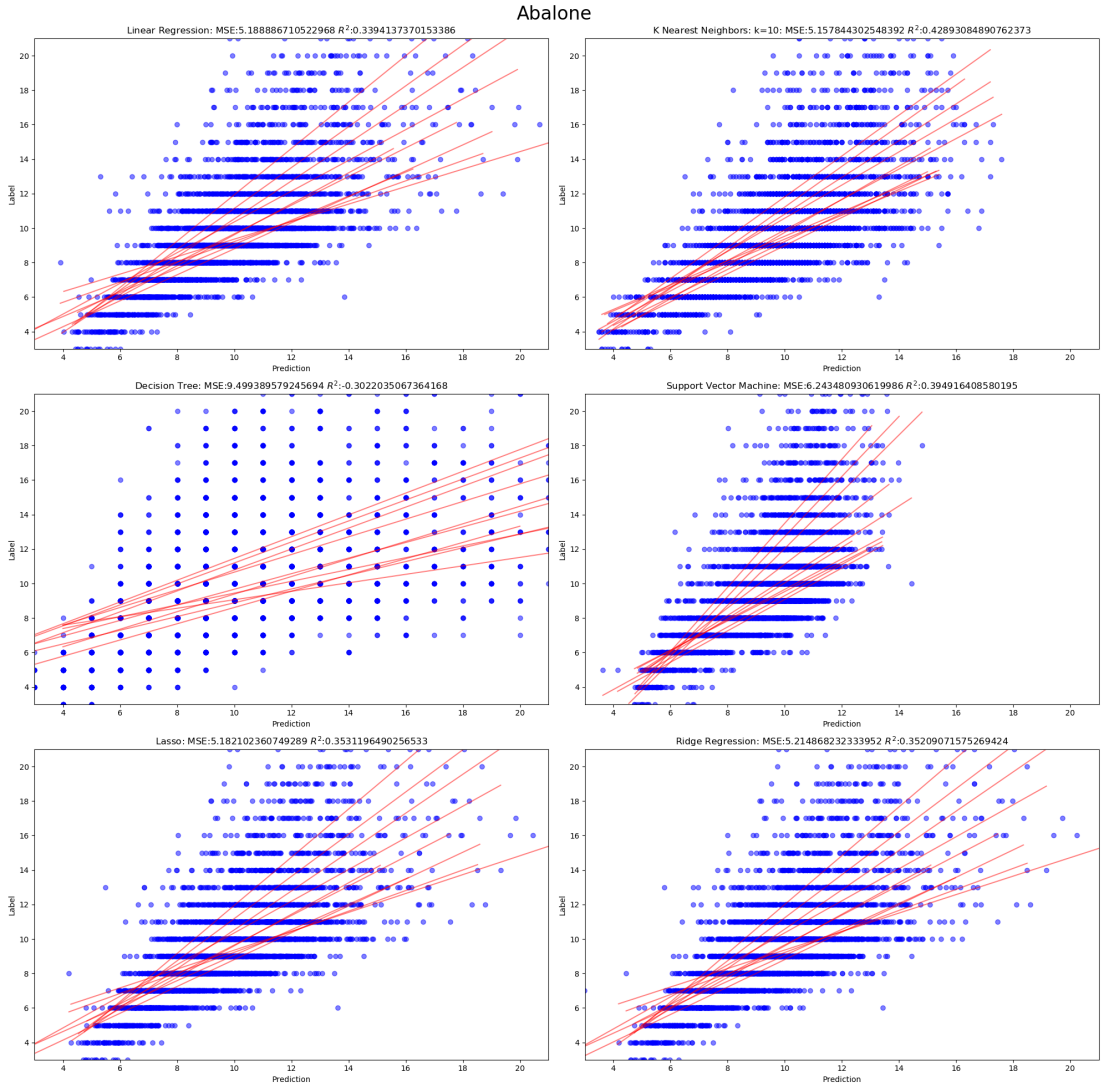
Wine Quality

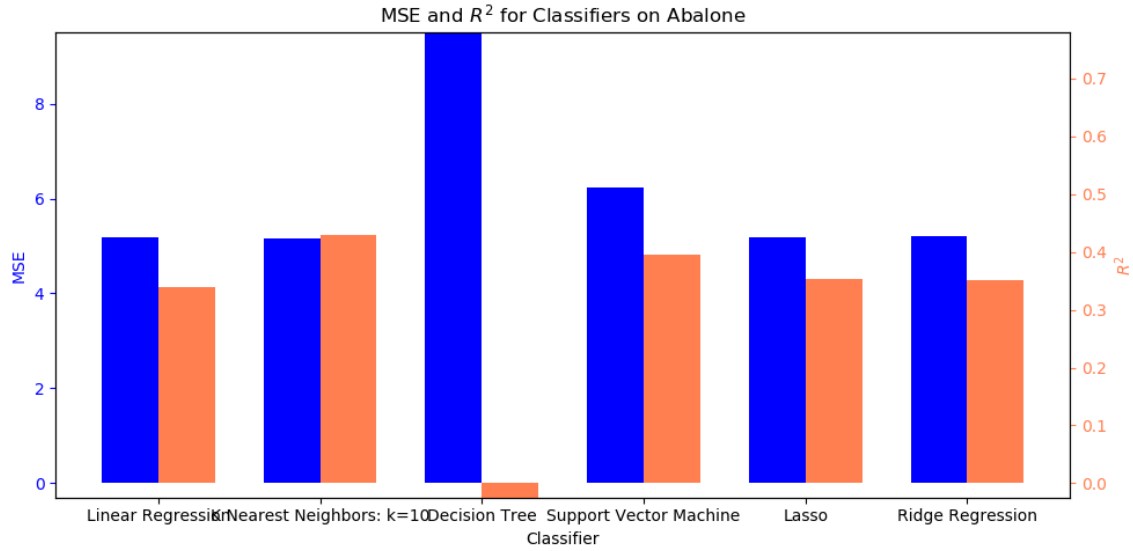




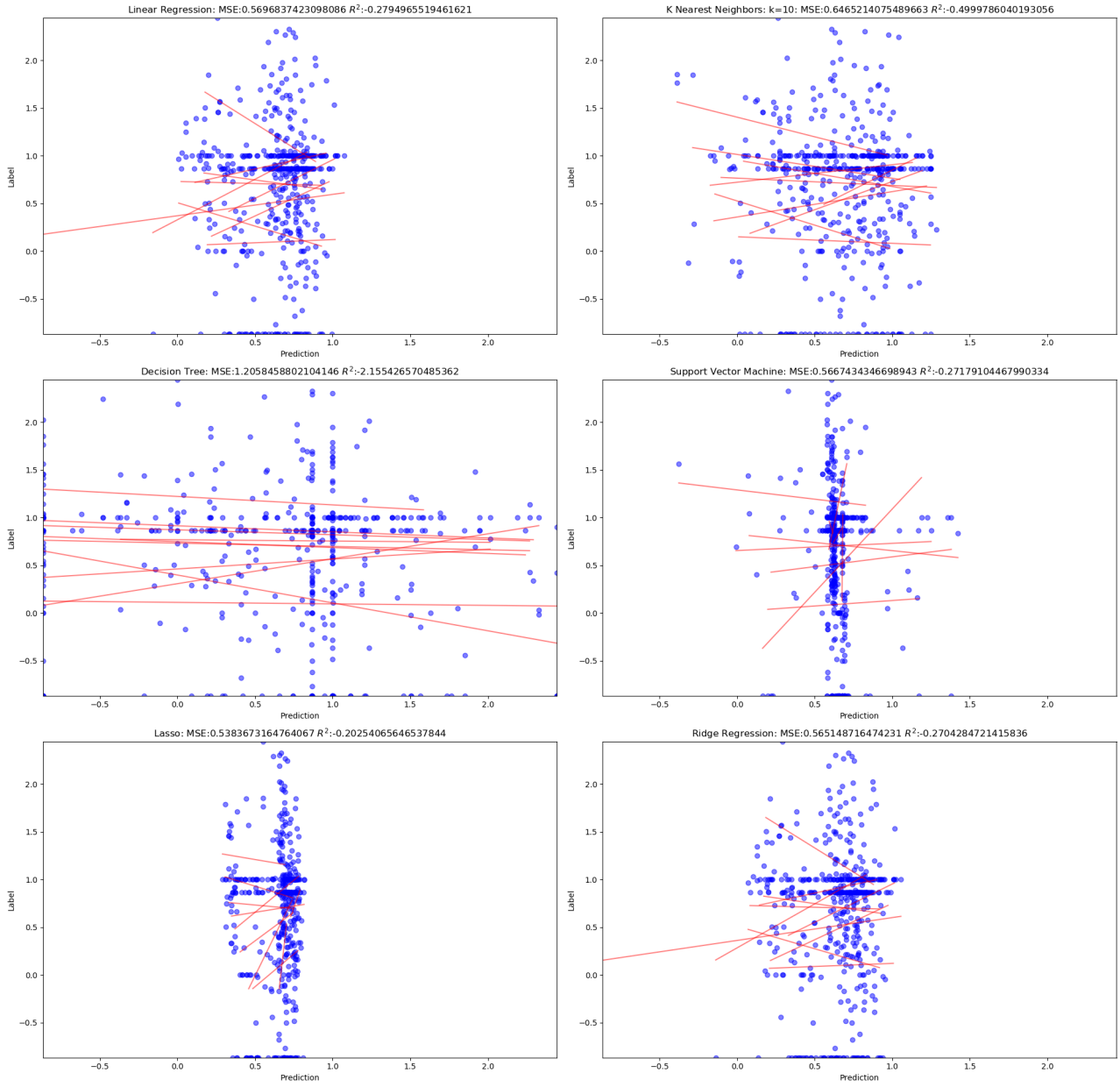
Update November 8, 2018

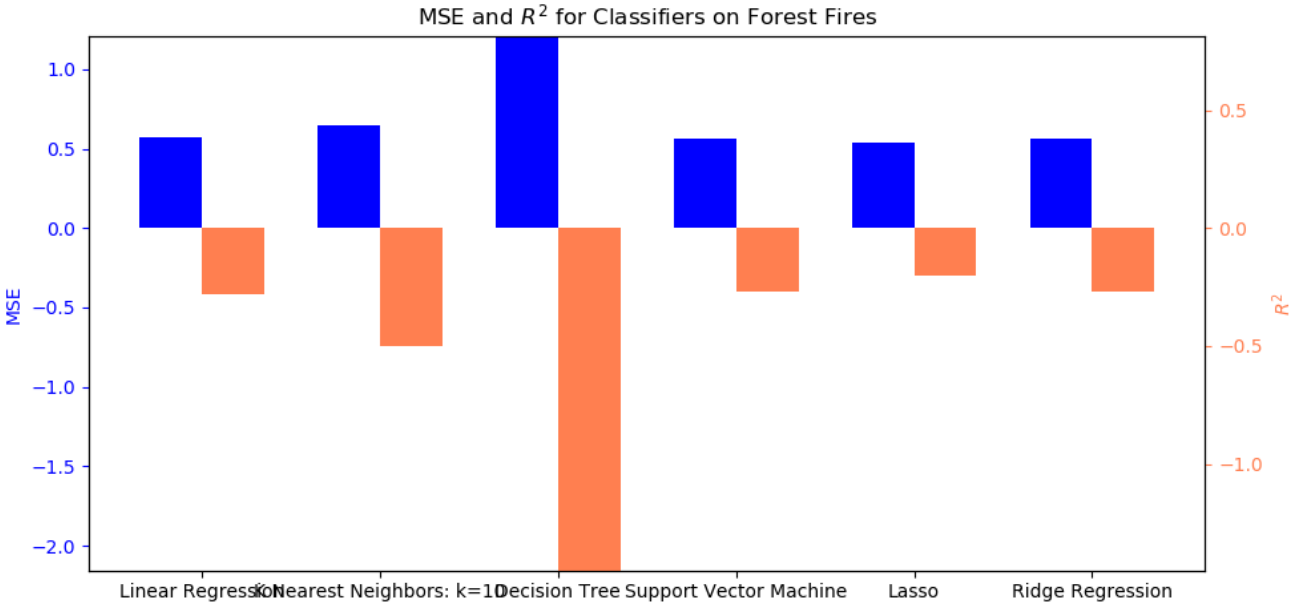
Lasso and Ridge Regression were added the regressors:





Forest Fires





Wine Qualtiy