Machine Learning - Practical 1

Handin 1: Bar chart

Distribution of wine quality (= y values) in White wine data set.
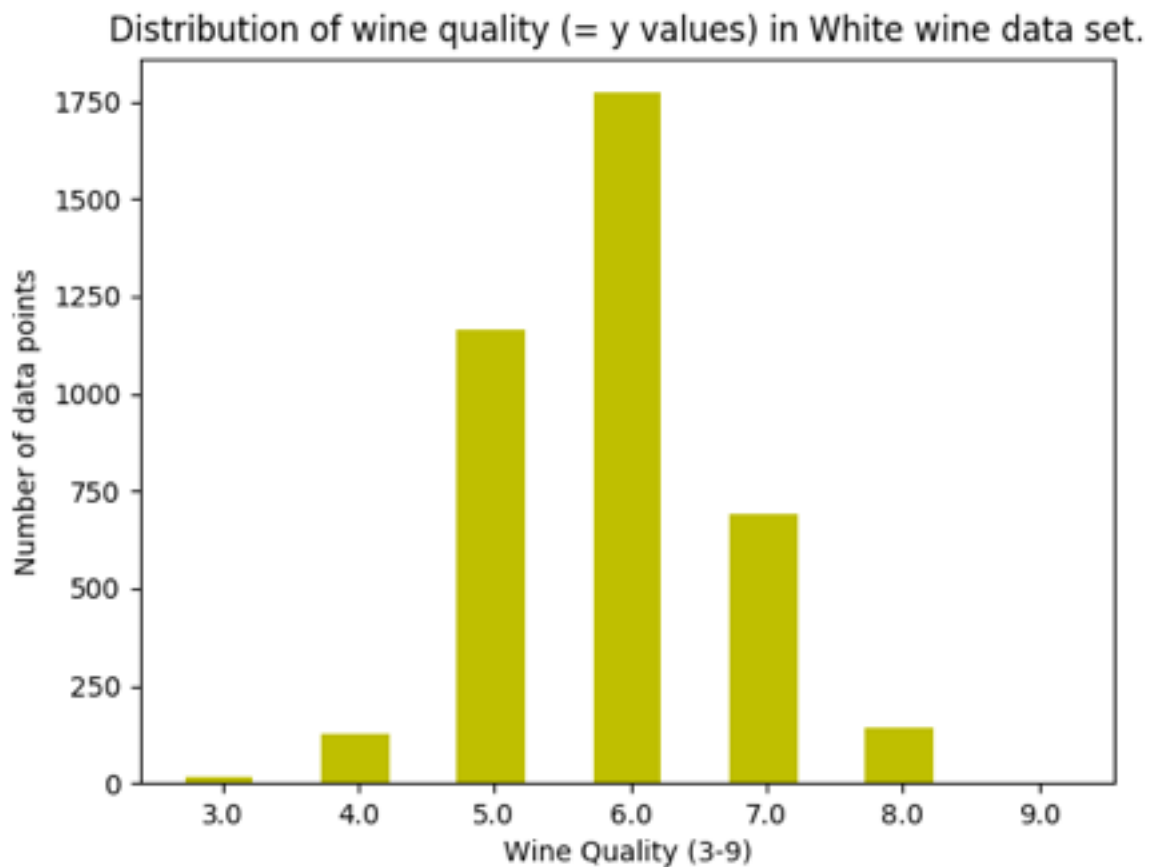
Handin 2: MSE of Zero R
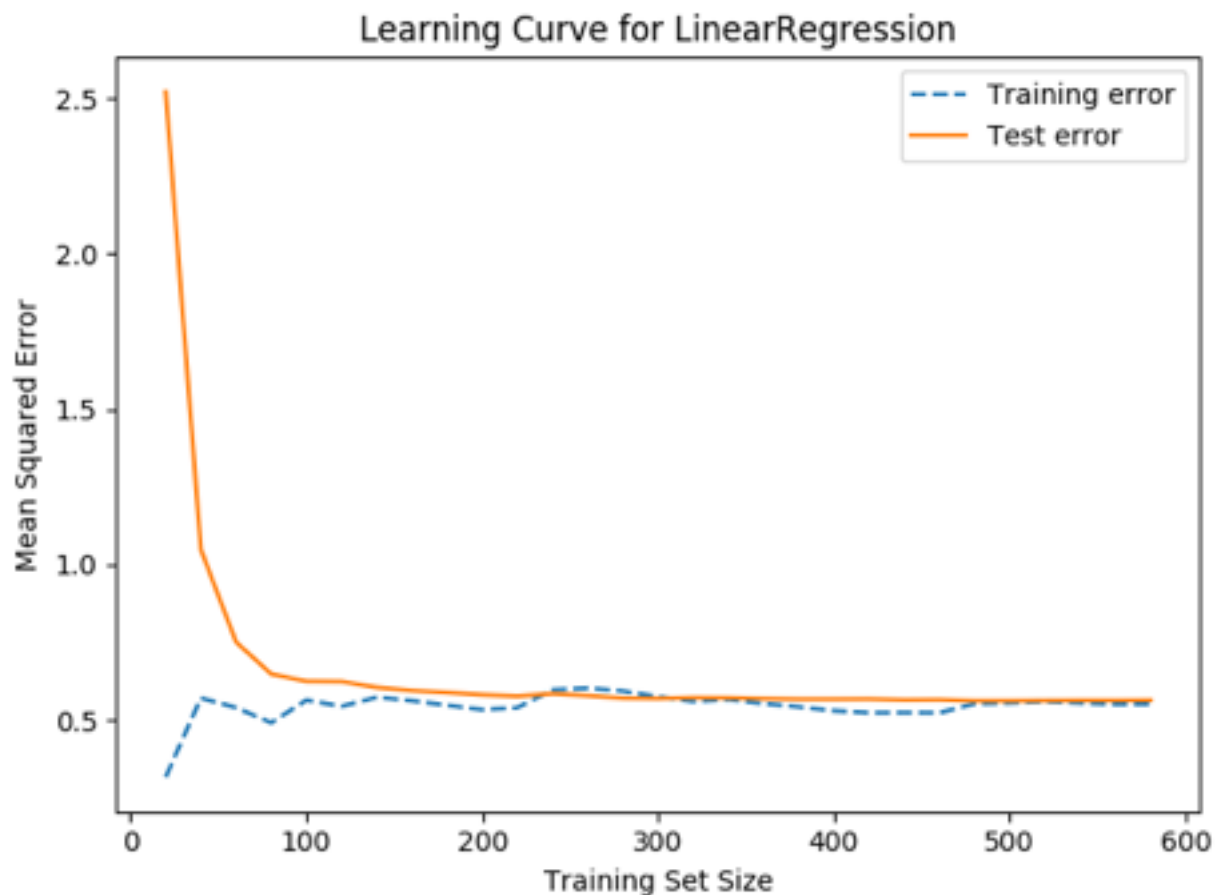
```
MSE of ZeroR on trainig set: 0.7768
MSE of ZeroR on trainig set: 0.8139
```

Handin 3: MSE of Linear Model using Least Squares

```
MSE of LinReg on trainig set: 0.5640
MSE of LinReg on test set: 0.5607
```

Handin 4: Learning curves for Linear Model

Conclusion: The model soon stops overfitting (see increase of training error when size of training set grows from 20 to 40) and rapidly improves on the test error. It's hard to tell if the model is underfitting or perfectly fine, as we have no further knowledge of the underlying distribution (i.e., if it is perfectly linear, how much noise was added, …) and therefore cannot judge just how well our model is fitting the data.