Raw_Me

12/11/2017

CSCI-3656

Homework 5

Q1:

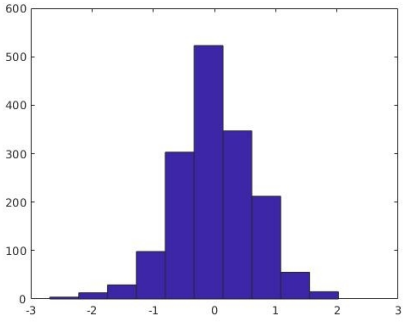
I used this Matlab code:

```
[y, X, labels] = process_datafile('winequality-red.csv');
e = ones(size(X,1), 1);
[b, bint, r, rint, stats] = regress(y,[e X]);
```

Q2:

I used this Matlab code:

```
hist(r)
```



It looks like normally distributed very close to a bell curve. (not the optimal).

Q3:

[b,bint,r,rint,stats] = regress(y,X) returns a 1-by-4 vector stats that contains, in order, the R^2 statistic, the F statistic and its p value, and an estimate of the error variance.

R^2 = 0.3606

which is not close to 1 as this R^2 value which indicates that about 36% of the model explains all the variability of the response data around its mean. In general, the model does not fit the good enough.

Q4:

I used this Matlab code:

```
std(b).*(std(X)/std(y))
```

fixed acidity	18.4751
volatile acidity	1.9000
citric acid	2.0671
residual sugar	14.9610
chlorides	0.4994
free sulfur dioxide	110.9946
total sulfur dioxide	349.0582
density	0.0200
pH	1.6382
sulphates	1.7987
alcohol	11.3080

The three most important are free sulfur dioxide, total sulfur dioxide, and fixed acidity.

Q5:

I used this matlab code:

```
pre = [1, 5.1 , 0.3 , 0.8 , 11.0 , 0.2 , 68.5 , 15.7 , 1.0 , 3.2 , 1.6 , 13.6 ];

score = pre*b;
```

score = 7.6914