Project 2

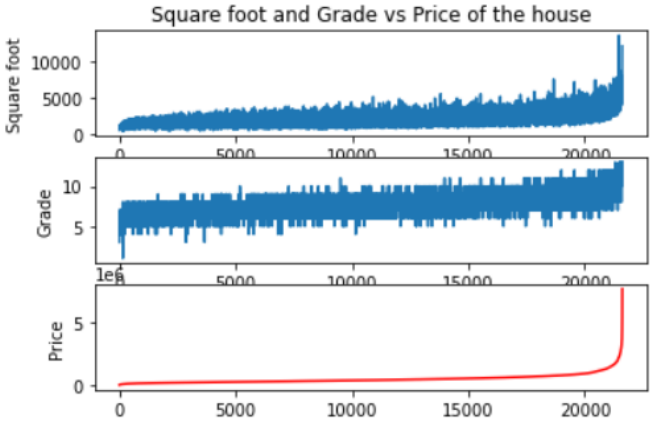
Jeet Shah (G41262537)

Tools and language used: Python, SciKit, numpy, panda, seaborn, matplot

Question 1:

Here we are given a dataset that contains house sale prices for King County, which includes Seattle. It includes homes sold between May 2014 and May 2015. There are many features and depending on the feature and it's value, house price varies.

The most correlated feature from all the given features is sqft_living(it can be seen in the heatmap). Hence, I have used it as input from all other features.



The result is shown below.

```
Epoch Number:  2400
Coefficient is:  258512.04450724574
Cost is:  68529308801.34894
Intercept is:  538550.535517933
```

```
Mean Absolute Error:  173407.8459421428
```

```
R2 Score:  0.48702424474467476
Time taken: 0.24176454544067383
```

Link: https://github.com/jeetsj/ML_GWU/blob/master/Ques_1.ipynb

Question 2:

Here, sample entropy, information gain and decision tree is asked for a given set of training examples. Please, go to the link below to see the solution.

Link: https://github.com/jeetsj/ML_GWU/blob/master/Ques_2.pdf

Question 3:

We are asked to implement a simple Perceptron classifier to classify digits. Please, go to the link below to see the solution.

Link: https://github.com/jeetsj/ML_GWU/blob/master/python_HW2_Perceptron.ipynb

Question 4:

Here, we are given data of breast cancer diagnosis. There are various features. The cancer is diagnosed in two class, class 2 – Benign and class 4 – Malign.

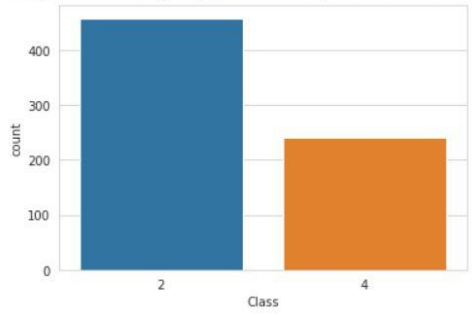
There is one feature “Bare Nuclei”, it has some missing values. I have updated those value by finding average of the values of that feature.

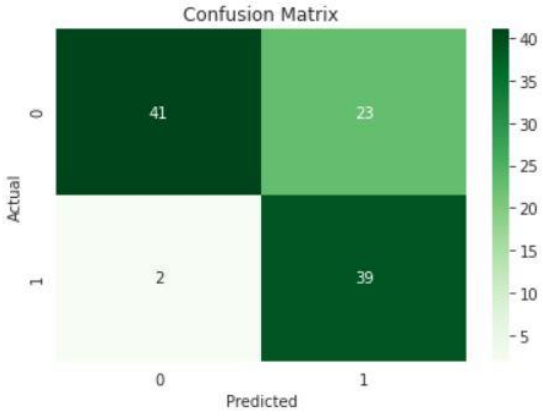
The SVM will learn using the stochastic gradient descent algorithm. SGD minimizes a function by following the gradients of the cost function. For regulizing parameter, I have chosen 1/epoch, so this parameter will decrease, as the number of epochs increases.

The accuracy of the algorithm is around 83%.

```
# Class = 4 is Malignant and class 2 is Benign.
```

```
/usr/local/lib/python3.6/dist-packages/seaborn/_decorators.p
Pass the following variable as a keyword arg: x. From versio
<matplotlib.axes._subplots.AxesSubplot at 0x7f64dae86240>
```





Link: https://github.com/jeetsj/ML_GWU/blob/master/Ques4_SVM.ipynb