CSE250 HW4 Q5

October 30, 2018

```
In [2]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        from numpy.linalg import *
        import math
```

```
In [3]: data00 = pd.read_csv('hw4_nasdaq00.txt', sep=" ", header=None)
        data01 = pd.read_csv('hw4_nasdaq01.txt', sep=" ", header=None)
```

0.0.1 (a)

```
In [106]: def transform_data(df):
              df_x1 = df.values[2:df.shape[0]-1]
              df_x2 = df.values[1:df.shape[0]-2]
              df_x3 = df.values[0:df.shape[0]-3]
              result = np.concatenate((df_x1, df_x2, df_x3), axis=1)
              return result
```

```
In [107]: df00 = transform_data(data00)
          df01 = transform_data(data01)
```

```
In [108]: def ComputeA(df):
              A = np.zeros((3,3))
              for i in range(df.shape[0]):
                  outer_product = np.outer(df[i,],df[i,])
                  A += outer_product
              return A

          def ComputeB(data,df):
              b = np.ones((1, 3))
              for i in range(df.shape[0]):
                  b += data.values[i+3,0]*df[i,]
              return b
```

```
In [109]: A = ComputeA(df00)
          b = np.transpose(ComputeB(data00,df00))
```

```
In [121]: weight = np.linalg.solve(A, b)
          a_1 = weight[0]
          a_2 = weight[1]
          a_3 = weight[2]
          print("a_1 = ", a_1)
          print("a_2 = ", a_2)
          print("a_3 = ", a_3)
```

```
a_1 =  [0.95067337]
a_2 =  [0.01560133]
a_3 =  [0.03189568]
```

0.0.2 (b)

```
In [119]: def ComputeMSE(weight,df,data):
              predict = np.dot(df,weight)
              SE = (data.values[3:,] - predict)**2
              MSE = np.mean(SE)
              return MSE

          MSE_train = ComputeMSE(weight,df00,data00)
          MSE_test = ComputeMSE(weight,df01,data01)
          print("MSE of training", MSE_train)
          print("MSE of testing", MSE_test)
```

```
MSE of training 13902.40107636789
MSE of testing 2985.0979239598764
```

The MSE of testing is smaller than training, and RMSE of 54.64 is reasonable for stock market prediction, where the prices are between 1000 and 10000. Therefore, I would recommend this linear model.