hw4prob4

October 28, 2021

1 Problem 4 Homework 4 CSE 250A Probabilistic Reasoning & Learning

a)

How accurately can the index on one day be predicted by a linear combination of the three preceding indices? Using only data from the year 2000 , compute the linear coefficients (a_1, a_2, a_3) that maximize the log-likelihood $\mathcal{L} = \sum_t \log P(x_t|x_{t-1}, x_{t-2}, x_{t-3})$, where:

$$P(x_t|x_{t-1}, x_{t-2}, x_{t-3}) = \frac{1}{\sqrt{2\pi}}\exp[-\frac{1}{2}(x_t - a_1x_{t-1} - a_2x_{t-2} - a_3x_{t-3})^2]$$

and the sum is over business days in the year 2000 (starting from the fourth day).

[1]:
```
import numpy as np
from numpy.linalg import inv
from numpy import transpose as T
```

[2]:
```
# Initializating the matrices for X and Y
X_2000 = []
Y_2000 = []

# Opening file and iterating through lines to predict
with open('nasdaq00.txt','r') as f:
    t_3 = -1
    t_2 = -1
    t_1 = -1
    for line in f.readlines():
        line = line.rstrip()
        if t_1 != -1:
            X_2000.append([t_3, t_2, t_1])
            Y_2000.append(float(line))
        t_1 = t_2
        t_2 = t_3
        t_3 = float(line)

# Computing a values
X_2000 = np.matrix(X_2000)
Y_2000 = np.matrix(Y_2000).reshape(-1,1)
```

```
a_2000 = inv(T(X_2000) * X_2000) * T(X_2000) * Y_2000
a_1,a_2,a_3 = round(a_2000.item(0),6), round(a_2000.item(1),6), round(a_2000.
 ↪item(2),6)
print(f"A values are (a1,a2,a3) = ({a_1},{a_2},{a_3}).")
```

```
A values are (a1,a2,a3) = (0.950673,0.015601,0.031896).
```

[3]:
```
# Initializating the matrices for X and Y
X_2001 = []
Y_2001 = []

# Opening file and iterating through lines to predict
with open('nasdaq01.txt','r') as f:
    t_3 = -1
    t_2 = -1
    t_1 = -1
    for line in f.readlines():
        line = line.rstrip()
        if t_1 != -1:
            X_2001.append([t_3, t_2, t_1])
            Y_2001.append(float(line))
        t_1 = t_2
        t_2 = t_3
        t_3 = float(line)

# Computing matrices for MSE computing
X_2001 = np.matrix(X_2001)
Y_2001 = np.matrix(Y_2001).reshape(-1,1)
a_2001 = inv(T(X_2001) * X_2001) * T(X_2001) * Y_2001
```

[4]:
```
rmse_2000 = np.sqrt(np.square(Y_2000 - np.dot(X_2000,a_2000)).mean())
rmse_2001 = np.sqrt(np.square(Y_2001 - np.dot(X_2001,a_2001)).mean())
print(f"RMSE Year 2000: {rmse_2000}")
print(f"RMSE Year 2001: {rmse_2001}")
```

```
RMSE Year 2000: 117.90844361778286
RMSE Year 2001: 54.486240219686444
```

We observe that the RMSE is lower in 2001 than in 2000, so the prediction is better on paper.

However, the stock market is very hard to predict, so this might just have been a coincidence.