# Assignment 2

All results are rounded to four decimal places in the document, for all code you need to install traceback module.

## Question 1

* 1. the answer could be seen by run the code (Question1.1\_HSBC.py, Question1.1\_SPY.py) for both two stock, also you need to install panads and colorama module for running the code without error. And panads module is for the beautiful table and colorama is for the label color.
  2. default probability p\* for my stock HSBC is 0.5192, and for SPY is 0.5351
  3. the probability of seeing “-, -, -, +” as opposed to seeing

“-, -, -, -“ for my stock HSBC is k=1 0.3502; k=2 0.3088; k=3 0.3056, and for SPY is k=1 0.4457; k=2 0.5405; k=3 0.5

* 1. the probability of seeing “+, +, +, +” as opposed to seeing

“+, +, +, -“ for my stock HSBC is k=1 0.6923; k=2 0.7342; k=3 0.7882, and for SPY is k=1 0.6522; k=2 0.6259; k=3 0.6143

The code for question1.2-1.4 could be seen at Question1.2-4\_HSBC.py and Question1.2-4\_SPY.py for both stock

## Question 2

2.1 the predicted label for W = 2, 3, 4 could be seen by run the code Question2-5\_HSBC.py and Question2-4\_SPY.py. Also you need to delete the # before the print in the predict() function, the answer will be shown as 2018-01-02 +.

2.2 the accuracy (what percentage of true label did I predicted correctly) has shown below in the table

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| accuracy | HSBC positive | HSBC negative | SPY positive | SPY negative |
| 2 | 0.7598 | 0.2329 | 0.6491 | 0.3119 |
| 3 | 0.6417 | 0.3855 | 0.5088 | 0.4083 |
| 4 | 0.6850 | 0.3494 | 0.4316 | 0.5138 |

2.3 for HSBC stock W = 4 gave me the highest accuracy, for SPY stock W = 2 gave me the highest accuracy

## Question 3

3.1 the ensemble labels could be seen by run the code Question2-5\_HSBC.py and Question2-4\_SPY.py also you need to get rid of the comment, you can see where it is in the main function.

3.2 for my stock the correct percentage of labels in year 4 and 5 is 51.49%; and for the S&P-500 is 46.52%

3.3 the accuracy on predicting”-“ is same as W = 3 for both stock

3.4 the accuracy on predicting”+“ is same as W = 3 for both stock

### Question 4

4.1-6 could be seen by code Question2-5\_HSBC.py and Question2-4\_PSY.py

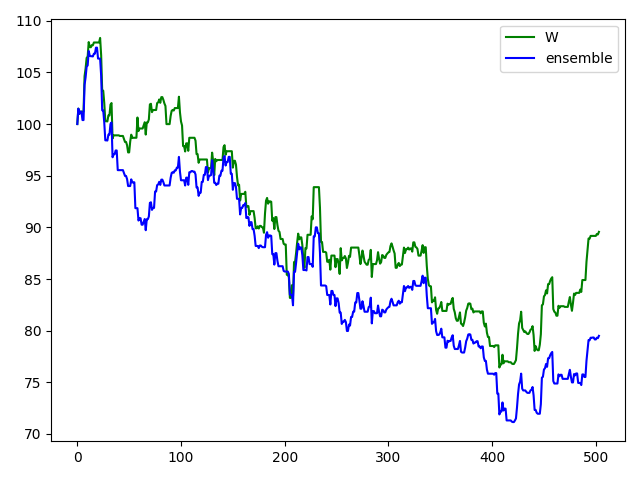
4.7 table

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| W | ticker | TP | FP | TN | FN | accuracy | TPR | TNR |
| 2 | SPY | 185 | 150 | 68 | 100 | 0.503 | 0.6491 | 0.3119 |
| 3 | SPY | 145 | 129 | 89 | 140 | 0.4652 | 0.5088 | 0.4083 |
| 4 | SPY | 123 | 106 | 112 | 162 | 0.4672 | 0.4316 | 0.5138 |
| ensemble | SPY | 145 | 129 | 89 | 140 | 0.4652 | 0.5088 | 0.4083 |
| 2 | HSBC | 193 | 191 | 58 | 61 | 0.499 | 0.7598 | 0.2329 |
| 3 | HSBC | 163 | 153 | 96 | 91 | 0.5149 | 0.6417 | 0.3855 |
| 4 | HSBC | 174 | 162 | 87 | 80 | 0.5189 | 0.685 | 0.3494 |
| ensemble | HSBC | 163 | 153 | 96 | 91 | 0.5149 | 0.6417 | 0.3855 |

4.8 At first, I thought that the predicted probability will increase when w grows, but this forecasting method only uses past data to predict, and does not consider the current market conditions that are closely related to stocks. From the prediction results alone, when w increases, the prediction of‘-’tends to be more accurate. For the high prediction rate of ‘+’ when w=2, just because most of the symbols are predicted to ‘+’. For the prediction result of ensemble learning, because the result of each day predicted by w=2, 3, 4 does not include‘-, +, -‘ or ‘+, -, +’. So the predicted label is the same as the label of w=3, so all the probabilities are absolutely the same

### Question 5

5.1 plot



5.2 the green line is shown based on the best W\* and blue line is shown based on ensemble. Also, I used the real data to draw a graph, the final price is between my best W and ensemble. Year 5 is more accuracy than year 4, and due to W=4 predict more negative correct, so the price of W=4 is more higher than ensemble.

If you have any question about the code please email me thank you so much.