Assignment 2

**Responses tagged <AA>**

**I have chosen JNPR ticker instead of a ticker based on my last name because i have invested in JNPR ticker and i’d like to use this exercise to study my investments. Right now, i’m on a buy and hold strategy.**

**Question 1:** You have a csv table of daily returns for your stosk and for S&P-500 (”spy” ticker).

1. For each file, read them into a pandas frame and add a column ”True Label”.

In that column, for each day (row) i with daily return ri ≥ 0 you assign a ” + ” label (”up day”). For each day i with daily return ri < 0 you assign ” − ” (”down days”). You do this for every day for all 5 years both both tickers. For example, if your initial dataframe were

you will add an additional column ”True Label” and have data as shown in Table 2. Your daily ”true labels” sequence is +, −, +, · · · +, −.

<AA>

**%%%%%%JNPR Ticker**

Date Year Month Day ... Return Short\_MA Long\_MA True Label

0 2016-01-04 2016 1 4 ... 0.000000 22.440000 22.4400 +

1 2016-01-05 2016 1 5 ... 0.005108 22.500000 22.5000 +

2 2016-01-06 2016 1 6 ... -0.027949 22.310000 22.3100 -

3 2016-01-07 2016 1 7 ... -0.025766 22.072500 22.0725 -

4 2016-01-08 2016 1 8 ... -0.014182 21.870000 21.8700 -

... ... ... ... ... ... ... ... ... ...

1253 2020-12-23 2020 12 23 ... -0.001341 20.498571 19.9704 -

1254 2020-12-24 2020 12 24 ... 0.002686 20.526429 19.9724 +

1255 2020-12-28 2020 12 28 ... 0.007143 20.556429 19.9774 +

1256 2020-12-29 2020 12 29 ... -0.016401 20.585714 19.9766 -

1257 2020-12-30 2020 12 30 ... 0.000451 20.605000 19.9826 +

[1258 rows x 17 columns]

**%%%%%%SPY Ticker**

Date Year Month Day ... Return Short\_MA Long\_MA True Label

0 2016-01-04 2016 1 4 ... 0.000000 175.290000 175.2900 +

1 2016-01-05 2016 1 5 ... 0.001691 175.440000 175.4400 +

2 2016-01-06 2016 1 6 ... -0.012614 174.750000 174.7500 -

3 2016-01-07 2016 1 7 ... -0.023991 173.365000 173.3650 -

4 2016-01-08 2016 1 8 ... -0.010977 172.164000 172.1640 -

... ... ... ... ... ... ... ... ... ...

1253 2020-12-23 2020 12 23 ... 0.000898 352.741429 339.9586 +

1254 2020-12-24 2020 12 24 ... 0.003890 352.790714 340.3918 +

1255 2020-12-28 2020 12 28 ... 0.008591 353.109286 340.8940 +

1256 2020-12-29 2020 12 29 ... -0.001908 353.305714 341.3866 -

1257 2020-12-30 2020 12 30 ... 0.001427 353.765000 341.9904 +

[1258 rows x 17 columns]

1. take years 1,2 and 3. Let L be the number of trading days. Assuming 250 trading days per year, L will contain about 750 days. Let L− be all trading days with − labels and let L+ be all trading days with + labels. Assuming that all days are independent of each other and that the ratio of ”up” and ”down” days remains the same in the future, compute the default probability p∗ that the next day is a ”up” day.

<AA>

Probability is calculated by # of favourable events/ total events. Here # of favourable events is total number of up days. Total events is total days including both up and down days.

*Default probability of next day up for JNPR is (in percentage): 55.03978779840849*

*Default probability of next day up for SPY is (in percentage): 55.437665782493376*

1. take years 1, 2 and 3 What is the probability that after seeing k consecutive ”down days”, the next day is an ”up day”? For example, if k = 3, what is the probability of see- ing ”−, −, −, +” as opposed to seeing ”−, −, −, −”. Com- pute this for k = 1, 2, 3.

<AA> **JNPR**

Probability of K=1 pattern (-+) between 2016 to 2018 is (in percentage) 56.93

Probability of K=2 pattern (--+) between 2016 to 2018 is (in percentage) 24.48

Probability of K=3 pattern (---+) between 2016 to 2018 is (in percentage)10.32

**SPY**

Probability of K=1 pattern (-+) between 2016 to 2018 is (in percentage) 59.52

Probability of K=2 pattern (--+) between 2016 to 2018 is (in percentage) 24.11

Probability of K=3 pattern (---+) between 2016 to 2018 is (in percentage) 10.42

1. take years 1, 2 and 3. What is the probability that after seeing k consecutive ”up days”, the next day is still an ”up day”? For example, if k = 3, what is the probability of seeing ”+, +, +, +” as opposed to seeing ”+, +, +, −”? Compute this for k = 1, 2, 3.

<AA> **JNPR**

Probability of K=1 pattern (++) between 2016 to 2018 is (in percentage) 53.25

Probability of K=2 pattern (+++) between 2016 to 2018 is (in percentage) 27.47

Probability of K=3 pattern (++++) between 2016 to 2018 is (in percentage) 15.66

**SPY**

Probability of K=1 pattern (++) between 2016 to 2018 is (in percentage) 52.15

Probability of K=2 pattern (+++) between 2016 to 2018 is (in percentage) 26.08

Probability of K=3 pattern (++++) between 2016 to 2018 is (in percentage) 12.2

**Question 2:**

1. for W = 2, 3, 4, compute predicted labels for each day in year 4 and 5 based on true labels in years 1,2 and 3 only. Perform this for your ticker and for ”spy”.

<AA>

###### JNPR Ticker ######

###### W = 4 ######

Date Year Month ... Long\_MA True Label Predicted True Label

754 2019-01-02 2019 1 ... 24.3114 -

755 2019-01-03 2019 1 ... 24.2728 - +

756 2019-01-04 2019 1 ... 24.2502 + +

757 2019-01-07 2019 1 ... 24.2520 + -

758 2019-01-08 2019 1 ... 24.2414 + +

[5 rows x 18 columns]

###### W = 3 ######

Date Year Month ... Long\_MA True Label Predicted True Label

754 2019-01-02 2019 1 ... 24.3114 -

755 2019-01-03 2019 1 ... 24.2728 - +

756 2019-01-04 2019 1 ... 24.2502 + +

757 2019-01-07 2019 1 ... 24.2520 + -

758 2019-01-08 2019 1 ... 24.2414 + +

[5 rows x 18 columns]

###### W = 2 ######

Date Year Month ... Long\_MA True Label Predicted True Label

754 2019-01-02 2019 1 ... 24.3114 -

755 2019-01-03 2019 1 ... 24.2728 - +

756 2019-01-04 2019 1 ... 24.2502 + +

757 2019-01-07 2019 1 ... 24.2520 + -

758 2019-01-08 2019 1 ... 24.2414 + +

[5 rows x 18 columns]

###### SPY Ticker ######

###### W = 4 ######

Date Year Month ... Long\_MA True Label Predicted True Label

754 2019-01-02 2019 1 ... 244.3832 +

755 2019-01-03 2019 1 ... 243.8200 - +

756 2019-01-04 2019 1 ... 243.4310 + +

757 2019-01-07 2019 1 ... 243.1046 + +

758 2019-01-08 2019 1 ... 242.9748 + +

[5 rows x 18 columns]

Predicted accuracy for W = 4 is: 58.730000000000004

###### W = 3 ######

Date Year Month ... Long\_MA True Label Predicted True Label

754 2019-01-02 2019 1 ... 244.3832 +

755 2019-01-03 2019 1 ... 243.8200 - +

756 2019-01-04 2019 1 ... 243.4310 + +

757 2019-01-07 2019 1 ... 243.1046 + -

758 2019-01-08 2019 1 ... 242.9748 + +

[5 rows x 18 columns]

Predicted accuracy for W = 3 is: 51.980000000000004

###### W = 2 ######

Date Year Month ... Long\_MA True Label Predicted True Label

754 2019-01-02 2019 1 ... 244.3832 +

755 2019-01-03 2019 1 ... 243.8200 - +

756 2019-01-04 2019 1 ... 243.4310 + +

757 2019-01-07 2019 1 ... 243.1046 + -

758 2019-01-08 2019 1 ... 242.9748 + +

[5 rows x 18 columns]

Predicted accuracy for W = 2 is: 56.75

Process finished with exit code 0

2. for each W = 2, 3, 4, compute the accuracy - what percentage of true labels (both positive and negative) have you predicted correctly for the last two years.

<AA>

**<JNPR>**

Predicted accuracy for W = 4 is: 52.78

Predicted accuracy for W = 3 is: 51.980000000000004

Predicted accuracy for W = 2 is: 52.38

**<SPY>**

Predicted accuracy for W = 4 is: 58.730000000000004

Predicted accuracy for W = 3 is: 51.980000000000004

Predicted accuracy for W = 2 is: 56.75

3. which W ∗ value gave you the highest accuracy for your stock and and which W ∗ valuegave you the highest accuracy for S&P-500?

<AA> For JNPR and SPY the max accuracy was seen with W=4 data set.

**Question 3:**

1. compute ensemble labels for year 4 and 5 for both your stock and S&P-500.

<AA>

<JNPR>

###### Ensemble Label ######

Date Year Month ... True Label Predicted True Label Ensemble Label

754 2019-01-02 2019 1 ... - NaN

755 2019-01-03 2019 1 ... - + +

756 2019-01-04 2019 1 ... + + +

757 2019-01-07 2019 1 ... + - -

758 2019-01-08 2019 1 ... + + +

<SPY>

###### Ensemble Label ######

Date Year Month ... True Label Predicted True Label Ensemble Label

754 2019-01-02 2019 1 ... + NaN

755 2019-01-03 2019 1 ... - + +

756 2019-01-04 2019 1 ... + + +

757 2019-01-07 2019 1 ... + - -

758 2019-01-08 2019 1 ... + + +

2. for both S&P-500 and your ticker, what percentage of labels in year 4 and 5 do you compute correctly by using ensemble?

<AA> JNPR

Predicted accuracy for Ensemble Label for JNPR is: 52.38

SPY

Predicted accuracy for Ensemble Label for SPY is: 56.75

3. did you improve your accuracy on predicting ”−” labels by using ensemble compared to W = 2, 3, 4?

<AA> W=4 had better accuracy than ensemble. This is likely due to the distribution of labels across the years for a given day. i.e more data sets improved the accuracy W=4 is 4 data sets vs Ensemble is 3 data sets i.e 2016 to 2019

4. did you improve your accuracy on predicting ”+” labels by using ensemble compared to W = 2, 3, 4?

<AA> W=4 had better accuracy than ensemble. This is likely due to the distribution of labels across the years for a given day. i.e more data sets improved the accuracy W=4 is 4 data sets vs Ensemble is 3 data sets i.e 2016 to 2019

**Question 4:**

For W = 2, 3, 4 and ensemble, compute the following (both for your ticker and ”spy”) statistics based on years 4 and 5:

Answers for Q1-6 are computed in the code and results are summarized below.

<AA> Here is the summary table for both tickers.

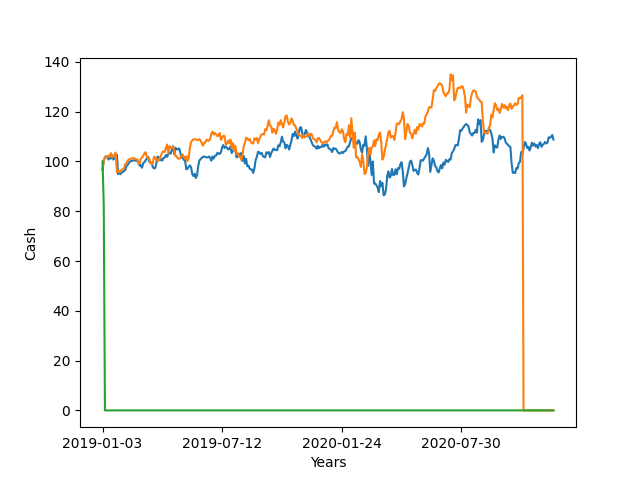
| W | Ticker | TP | FP | TN | FN | accuracy | TPR | TNR |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 2  3  4  ensemble | JNPR  JNPR  JNPR  JNPR | 102  82  101  102 | 90  71  88  90 | 30  49  32  30 | 30  50  31  30 | 52.38  51.98  52.78  52.38 | 77.2  62.1  76.5  77.2 | 25  40.8  26.6  25 |
| 2  3  4  ensemble | SPY  SPY  SPY  SPY | 123  88  123  123 | 82  59  77  82 | 20  43  25  20 | 27  62  27  27 | 56.75  51.98  58.7  56.75 | 82  58.6  82  82 | 19.6  42.1  24.5  19.6 |

**Question 5:**

At the beginning of year 4 you start with $100 dollars and trade for 2 years based on predicted labels.

1. take your stock. Plot the growth of your amount for 2 years if you trade based on best W ∗ and on ensemble. On the same graph, plot the growth of your portfolio for ”buy-and- hold” strategy

<AA> Graph for



W=4 is the blue line.

Ensemble is the orange line.

Red (starting at 0 with a vertical line ) is the buy and hold strategy.

2. examine your chart. Any patterns? (e.g any differences in year 4 and year 5)

<AA> W4 did peak for a brief period between 2019-03 to 2019-07, for rest of the period until 2020, the Ensemble had better returns.

The buy and hold strategy had lower returns of around 82$ because of not realizing the positive days.