

Question 1

1. read them into a pandas frame and add Predicting Daily Trading Labels column "True Label".

My stock table:

	Date	Year	Month	Day	Weekday	Week_Number	Year_Week	Open	High	Low	Close	Volume	Adj Close	Return	Short_MA	Long_MA	True Label
0	2016-01-04	2016	1	4	Monday	1	2016-01	475.85	477.84	466.50	472.77	70700	454.90	0.000000	454.900000	454.9000	+
1	2016-01-05	2016	1	5	Tuesday	1	2016-01	472.77	477.35	472.58	476.27	42600	458.27	0.007403	456.585000	456.5850	+
2	2016-01-06	2016	1	6	Wednesday	1	2016-01	471.25	474.91	469.81	473.22	44400	455.34	-0.006404	456.170000	456.1700	-
3	2016-01-07	2016	1	7	Thursday	1	2016-01	466.90	470.20	455.04	461.02	106700	443.60	-0.025781	453.027500	453.0275	-
4	2016-01-08	2016	1	8	Friday	1	2016-01	464.60	468.15	461.18	464.45	102400	446.90	0.007440	451.802000	451.8020	+
...
1253	2020-12-23	2020	12	23	Wednesday	51	2020-51	588.75	600.36	588.75	597.34	37200	597.34	0.024773	596.729286	588.6008	+
1254	2020-12-24	2020	12	24	Thursday	51	2020-51	594.09	599.12	591.49	598.15	32900	598.15	0.001356	595.825000	589.5798	+
1255	2020-12-28	2020	12	28	Monday	52	2020-52	601.12	607.97	599.14	599.14	39200	599.14	0.001655	595.611429	590.6136	+
1256	2020-12-29	2020	12	29	Tuesday	52	2020-52	602.00	604.07	594.75	598.28	54300	598.28	-0.001435	595.076429	591.6316	-
1257	2020-12-30	2020	12	30	Wednesday	52	2020-52	597.80	604.81	595.64	600.02	48100	600.02	0.002908	595.164286	592.7858	+

1258 rows × 17 columns

SPY stock table:

	Date	Year	Month	Day	Weekday	Week_Number	Year_Week	Open	High	Low	Close	Volume	Adj Close	Return	Short_MA	Long_MA	True Label
0	2016-01-04	2016	1	4	Monday	1	2016-01	200.49	201.03	198.59	201.02	222353500	178.05	0.000000	178.050000	178.0500	+
1	2016-01-05	2016	1	5	Tuesday	1	2016-01	201.40	201.90	200.05	201.36	110845800	178.35	0.001691	178.200000	178.2000	+
2	2016-01-06	2016	1	6	Wednesday	1	2016-01	198.34	200.06	197.60	198.82	152112600	176.10	-0.012614	177.500000	177.5000	-
3	2016-01-07	2016	1	7	Thursday	1	2016-01	195.33	197.44	193.59	194.05	213436100	171.88	-0.023992	176.095000	176.0950	-
4	2016-01-08	2016	1	8	Friday	1	2016-01	195.19	195.85	191.58	191.92	209817200	169.99	-0.010976	174.874000	174.8740	-
...
253	2020-12-23	2020	12	23	Wednesday	51	2020-51	368.28	369.62	367.22	367.57	46201400	358.58	0.000899	358.296429	345.3106	+
254	2020-12-24	2020	12	24	Thursday	51	2020-51	368.08	369.03	367.45	369.00	26457900	359.98	0.003890	358.346429	345.7506	+
255	2020-12-28	2020	12	28	Monday	52	2020-52	371.74	372.59	371.07	372.17	39000400	363.07	0.008591	358.670000	346.2608	+
256	2020-12-29	2020	12	29	Tuesday	52	2020-52	373.81	374.00	370.83	371.46	53680500	362.38	-0.001908	358.869286	346.7612	-
257	2020-12-30	2020	12	30	Wednesday	52	2020-52	372.34	373.10	371.57	371.99	49455300	362.89	0.001427	359.335714	347.3744	+

258 rows × 17 columns

- compute the default probability p^* that the next day is a "up" day.

The probability p that the next day is a "up" day of Y is 52.92%
 The probability p that the next day is a "up" day of SPY is 55.44%

- take years 1, 2 and 3 What is the probability that after seeing k consecutive "down days", the next day is an "up day"?

The probability of that after seeing 1 consecutive "down days" the next day is an "up day" of Y is 64.26%
 The probability of that after seeing 2 consecutive "down days" the next day is an "up day" of Y is 68.18%
 The probability of that after seeing 3 consecutive "down days" the next day is an "up day" of Y is 70.18%
 The probability of that after seeing 1 consecutive "down days" the next day is an "up day" of SPY is 67.57%
 The probability of that after seeing 2 consecutive "down days" the next day is an "up day" of SPY is 69.23%
 The probability of that after seeing 3 consecutive "down days" the next day is an "up day" of SPY is 71.43%

- 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"?

The probability of that after seeing 1 consecutive "up days" the next day is an "down day" of Y is 59.76%
 The probability of that after seeing 2 consecutive "up days" the next day is an "down day" of Y is 61.59%
 The probability of that after seeing 3 consecutive "up days" the next day is an "down day" of Y is 62.16%
 The probability of that after seeing 1 consecutive "up days" the next day is an "down day" of SPY is 58.31%
 The probability of that after seeing 2 consecutive "up days" the next day is an "down day" of SPY is 61.71%
 The probability of that after seeing 3 consecutive "up days" the next day is an "down day" of SPY is 69.05%

Question 2

- 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.

Table for my stock:

True Label	W2 Pred Label	W3 Pred Label	W4 Pred Label
+	+	+	-
-	-	+	+
+	+	+	+
+	+	+	-
+	-	+	+
...
+	+	+	+
+	+	+	+
+	-	+	+
-	-	-	-
+	+	+	+

Table for SPY stock:

True Label	W2 Pred Label	W3 Pred Label	W4 Pred Label
+	+	+	-
-	-	+	+
+	+	+	+
+	+	+	-
+	-	+	+
...
+	+	+	+
+	+	+	+
+	-	+	+
-	-	-	-
+	+	+	+

2. for each $W = 2, 3, 4$, compute the accuracy

The total accuracy for stock SPY W2 is 53.37%
 The accuracy on "-" for stock SPY W2 is 33.49%
 The accuracy on "+" for stock SPY W2 is 67.46%

The total accuracy for stock SPY W3 is 58.53%
 The accuracy on "-" for stock SPY W3 is 22.97%
 The accuracy on "+" for stock SPY W3 is 83.73%

The total accuracy for stock SPY W4 is 58.93%
 The accuracy on "-" for stock SPY W4 is 33.01%
 The accuracy on "+" for stock SPY W4 is 77.29%

The total accuracy for stock Y W2 is 50.79%
 The accuracy on "-" for stock Y W2 is 54.01%
 The accuracy on "+" for stock Y W2 is 47.94%

The total accuracy for stock Y W3 is 55.16%
 The accuracy on "-" for stock Y W3 is 31.65%
 The accuracy on "+" for stock Y W3 is 76.03%

The total accuracy for stock Y W4 is 57.94%
 The accuracy on "-" for stock Y W4 is 45.15%
 The accuracy on "+" for stock Y W4 is 69.29%

3. which W_* value gave you the highest accuracy for your stock and which W_* value gave you the highest accuracy for S&P-500?

Both for my stock Y and stock SPY, W4 gives the highest accuracy.

Question 3

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

Ensemble labels for year 4 and 5 for my stock:

True Label	W2 Pred Label	W3 Pred Label	W4 Pred Label	W2 validation	W3 validation	W4 validation	Ensemble Pred Label
-	+	-	-	0	0	0	-
-	-	+	-	0	0	0	-
+	+	+	+	1	1	1	+
-	+	+	-	0	0	0	+
-	-	+	+	0	0	0	+
...
+	+	+	+	1	1	1	+
+	+	+	+	1	1	1	+
+	-	+	+	0	1	1	+
-	-	-	+	0	0	0	-
+	-	+	+	0	1	1	+

Ensemble labels for year 4 and 5 for SPY stock:

True Label	W2 Pred Label	W3 Pred Label	W4 Pred Label	W2 validation	W3 validation	W4 validation	Ensemble Pred Label
+	+	+	-	1	1	0	+
-	-	+	+	0	0	0	+
+	+	+	+	1	1	1	+
+	+	+	-	1	1	0	+
+	-	+	+	0	1	1	+
...
+	+	+	+	1	1	1	+
+	+	+	+	1	1	1	+
+	-	+	+	0	1	1	+
-	-	-	-	0	0	0	-
+	+	+	+	1	1	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?

The accuracy of ensemble predicted for stock SPY is 58.53%
The accuracy for "-" of ensemble predicted for stock SPY is 22.97%
The accuracy for "+" of ensemble predicted for stock SPY is 83.73%

The accuracy of ensemble predicted for stock Y is 57.94%
The accuracy for "-" of ensemble predicted for stock Y is 47.68%
The accuracy for "+" of ensemble predicted for stock Y is 67.04%

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

My accuracy of ensemble for stock SPY is 22.97%, my accuracy of W=2,3,4 is 33.49%, 22.97%, 33.01% respectively

My accuracy of ensemble for stock Y is 47.68%, my accuracy of W=2,3,4 is 54.01%, 31.65%, 45.15% respectively

So in conclusion, both of my stock Y and SPY stock are not improved accuracy on predicting "-" labels by using ensemble compared to W = 2, 3, 4

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

My accuracy of ensemble for stock SPY is 83.73%, my accuracy of W=2,3,4 is 67.46%, 83.73%, 77.29% respectively

So the stock SPY not improved my accuracy on predicting "+" labels by using ensemble compared to W = 2, 3, 4

My accuracy of ensemble for stock Y is 67.04%, my accuracy of W=2,3,4 is 47.94%, 76.03%, 69.29% respectively

So my stock Y only improve my accuracy on predicting "+" labels by using ensemble compared to W = 2

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:

W	Ticker	TP	FP	TN	FN	Accuracy	TPR	TNR
2	S&P-500	199	139	70	96	53.37%	0.6746	0.3349
3	S&P-500	247	161	48	48	58.53%	0.8373	0.2297
4	S&P-500	228	140	69	67	58.93%	0.7729	0.3301
Ensemble	S&P-500	247	161	48	48	58.53%	0.8373	0.2297
2	Y	128	109	128	139	50.79%	0.4794	0.5401
3	Y	203	162	75	64	55.16%	0.7603	0.3165
4	Y	185	130	107	82	57.94%	0.6929	0.4515
Ensemble	Y	179	124	113	88	57.94%	0.6704	0.4768

Question 5

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

Trade for my stock Y:



From the picture, we can see easily that all the 3 methods have the same trend, which is slowly increasing, then a sharp drop in a short time, and finally get a slow and unsteady increase. By using the W4 prediction we can get the best profit if we use buy and hold strategy, we will get a very low income which is even smaller that the start money 100 dollars. It means that our W4 prediction and the thought that makes the prediction by using data are making sense.

Trade for the stock SPY:

