<u>**CS677 HW 2**</u>

By: Ahnaf Tajwar

Question 1

1. See code
2. Positive Probability
    a. SPY = 0.5410
    b. TSLA = 0.5027
3. Probability that for k consecutive ”down days”, the next day is an ”up day”
    a. SPY
        i. K = 1
            1. 0.5925
        ii. K = 2
            1. 0.6429
        iii. K = 3
            1. 0.6122
    b. TSLA
        i. K = 1
            1. 0.5267
        ii. K = 2
            1. 0.5650
        iii. K = 3
            1. 0.5714
4. Probability that for k consecutive ”up days”, the next day is an ”down day”
    a. SPY
        i. K = 1
            1. 0.5037
        ii. K = 2
            1. 0.4975
        iii. K = 3
            1. 0.5392
    b. TSLA
        i. K = 1
            1. 0.5224
        ii. K = 2
            1. 0.5028
        iii. K = 3
            1. 0.5778

Question 2

1. See code
2. Accuracy

a. SPY
    i. Overall Prediction Accuracy:
        1. {'**W2**': 0.5269461077844312, '**W3**': 0.532, '**W4**': 0.5370741482965932}
    ii. Positive Prediction Accuracy:
        1. {'**W2**': 0.7578947368421053, '**W3**': 0.6807017543859649, '**W4**': 0.6070175438596491}
    iii. Negative Prediction Accuracy:
        1. {'**W2**': 0.22018348623853212, '**W3**': 0.3302752293577982, '**W4**': 0.43577981651376146}
b. TSLA
    i. Overall Prediction Accuracy:
        1. {'**W2**': 0.4800796812749004, '**W3**': 0.500998003992016, '**W4**': 0.49}
    ii. Positive Prediction Accuracy:
        1. {'**W2**': 0.20216606498194944, '**W3**': 0.4548736462093863, '**W4**': 0.5234657039711191}
    iii. Negative Prediction Accuracy:
        1. {'**W2**': 0.8149779735682819, '**W3**': 0.5506607929515418, '**W4**': 0.44052863436123346}

3. Highest Accuracy
    a. SPY
        i. W = 4 gave me the highest overall accuracy for SPY
        ii. W = 2 gave me the highest positive accuracy for SPY
        iii. W = 4 gave me the highest negative accuracy for SPY
    b. TSLA
        i. W = 3 gave me the highest overall accuracy for TSLA
        ii. W = 4 gave me the highest positive accuracy for TSLA
        iii. W = 2 gave me the highest negative accuracy for TSLA

Question 3

1. See code
2. Percentage correct using ensemble
    a. SPY
        i. Percentage of correctly predicted ensembles: 0.5288270377733598
        ii. Percentage of correctly predicted positive ensembles: 0.6807017543859649
        iii. Percentage of correctly predicted negative ensembles: 0.3302752293577982
    b. TSLA
        i. Percentage of correctly predicted ensembles: 0.48214285714285715
        ii. Percentage of correctly predicted positive ensembles: 0.3898916967509025

iii. Percentage of correctly predicted negative ensembles: 0.5947136563876652

3. Accuracy improvement for negative labels
    a. SPY
        i. The accuracy for predicting negative labels using ensemble was not improved for W4, but was improved for W2, and remained the same for W3
    b. TSLA
        i. The accuracy for predicting negative labels using ensemble was not improved for W2, but was improved for W3 and W4
4. Accuracy improvement for positive labels
    a. SPY
        i. The accuracy for predicting positive labels using ensemble was not improved for W2, but was improved for W4, and remained the same for W3
    b. TSLA
        i. The accuracy for predicting positive labels using ensemble was not improved for W3 and W4, but was improved for W2

Question 4

1-6. Summarized in table below:

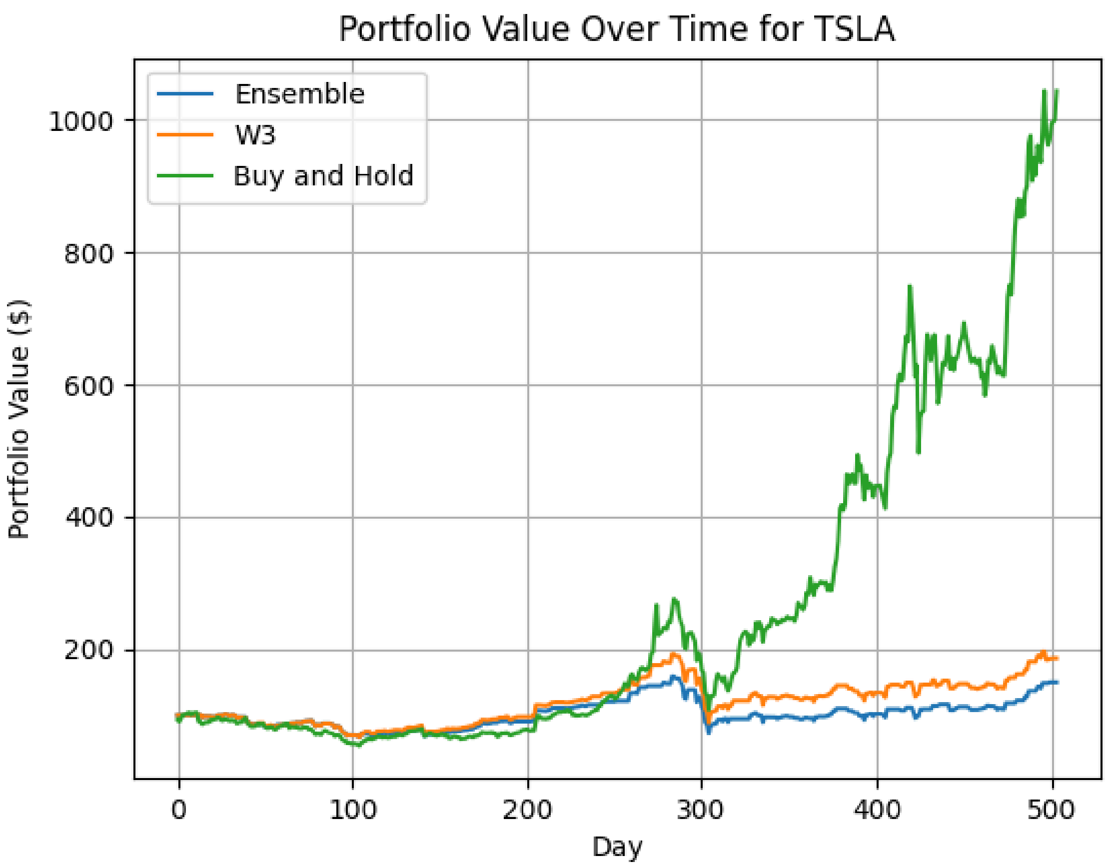
7. Statistics for SPY and TSLA

W	ticker	TP	FP	TN	FN	accuracy	TPR	TNR
W2	SPY.csv	216	170	48	67	0.526946107784431 2	0.763250883392226 1	0.220183486238532 12
W3	SPY.csv	194	146	72	88	0.532	0.687943262411347 5	0.330275229357798 2
W4	SPY.csv	173	123	95	108	0.537074148296593 2	0.615658362989323 9	0.435779816513761 46
Ensemble	SPY.csv	194	146	72	87	0.528827037773359 8	0.690391459074733 1	0.330275229357798 2
W2	TSLA.csv	56	40	185	221	0.480079681274900 4	0.202166064981949 44	0.822222222222222 2
W3	TSLA.csv	126	100	125	150	0.500998	0.456521739130434 76	0.555555555555555 6
W4	TSLA.csv	145	125	100	130	0.49	0.527272727272727 2	0.444444444444444 4

Ensemble	TSLA.csv	108	90	135	168	0.48214285714285715	0.391304	0.6

8. For SPY, the accuracy seems to be best for W4, while for TSLA it is W3. In addition, for SPY, the TPR is best for W2 and TNR is best for W4. For TSLA, the TPY is best for W4 and TNR is best for W2. The Ensemble method in these two cases did not perform better than the other methods, but was more like an average of all methods.

Question 5



1.
2. For year 4, it seems like the buy and hold strategy performed worse than both the W3 and Ensemble strategy. However, around the start of year 5 is when it greatly outperformed the other two, ending with over 5x portfolio compared to W3 and Ensemble. The W3 and Ensemble methods stays relatively similar throughout years 4 and 5, with the W3 method performing slightly better overall. For all methods, the portfolio had a sharp spike and then followed by a sharp dip near the beginning of year 5 (Day 250-300).