Divya Thomas

Class: CS 677

Date: 3/27/2023

Non-code answers to assignment.

**Question 1-1.** Pandas frame is printed in question1.py file.

**Question 1-2.** Please see calculated output below

TGT: The default probability that the next day after 2020 is an 'up' day is 0.5317460317460317

SPY: The default probability that the next day after 2020 is an 'up' day is 0.5701058201058201

**Question 1-3.**

TGT:

The probability of there being an 'up' day after 1 consecutive 'down' day(s) is 0.559

The probability of there being an 'up' day after 2 consecutive 'down' day(s) is 0.564

The probability of there being an 'up' day after 3 consecutive 'down' day(s) is 0.544

SPY:

The probability of there being an 'up' day after 1 consecutive 'down' day(s) is 0.578

The probability of there being an 'up' day after 2 consecutive 'down' day(s) is 0.540

The probability of there being an 'up' day after 3 consecutive 'down' day(s) is 0.524

**Question 1-4.**

TGT:

The probability of there being an 'up' day after 1 consecutive 'up' day(s) is 0.506

The probability of there being an 'up' day after 2 consecutive 'up' day(s) is 0.488

The probability of there being an 'up' day after 3 consecutive 'up' day(s) is 0.505

SPY:

The probability of there being an 'up' day after 1 consecutive 'up' day(s) is 0.563

The probability of there being an 'up' day after 2 consecutive 'up' day(s) is 0.568

The probability of there being an 'up' day after 3 consecutive 'up' day(s) is 0.474

**Question 2-1.** Please see output of W 2-4 values on question2.py. Any W\* value that required to check into training data to retrieve the search sequence was replaced with a NaN value.

**Question 2-2.**

TGT:

W2 computations were 47.90% accurate

W3 computations were 49.20% accurate

W4 computations were 43.69% accurate

SPY:

W2 computations were 50.70% accurate

W3 computations were 50.80% accurate

W4 computations were 52.30% accurate

**Question 2-3.** According to the above accuracy information, please see my observations below.

For the TGT stock, W3 predictions were deemed to be the most accurate, even though all W predictions fell under 50% accuracy.

For the SPY stock, on the other hand, W4 predictions were around 2% more accurate than others.

**Question 3-1.** Please question3.py for dataframe with ensemble values for both TGT and SPY.

**Question 3-2.**

TGT: WE computations were 47.32% accurate

SPY: WE computations were 51.09% accurate

**Question 3-3.**

While predicting “-“ labels, accuracy was generally very low for all W values (less than 40%). Using ensemble actually worsened the accuracy of predicting these labels compared to most of the values of W 2-4. Ensemble accuracy was the 2nd least accurate out of the 4 for each ticker when it came to down days.

**Question 3-4.**

While predicting “+” labels, accuracy was fairly high for all W values, especially in comparison to down days. The accuracy of using ensemble in comparison to W 2-4 values was pretty much the same, but on the higher end. Ensemble accuracy was the 2nd most accurate out of the 4 for each ticker when it came to up days. The tendency to be more accurate for this one in comparison to down days could be due to the high default probability of both tickers for up days.

See comprehensive data below:

TGT:

W2 computations predicted 68.85% of up days and 25.10% of down days

W3 computations predicted 85.38% of up days and 10.70% of down days

W4 computations predicted 48.85% of up days and 38.27% of down days

WE computations predicted 76.54% of up days and 16.05% of down days

SPY:

W2 computations predicted 99.61% of up days and 0.00% of down days

W3 computations predicted 73.83% of up days and 27.13% of down days

W4 computations predicted 69.53% of up days and 34.41% of down days

WE computations predicted 81.25% of up days and 19.84% of down days

**Question 4-1.** See 4.7 table

**Question 4-2.** See 4.7 table

**Question 4-3.** See 4.7 table

**Question 4-4.** See 4.7 table

**Question 4-5.** See 4.7 table

**Question 4-6.** See 4.7 table

**Question 4-7.** Please see table below

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **W** | **Ticker** | **TP** | **FP** | **TN** | **FN** | **Accuracy (%)** | **TPR** | **TNR** |
| **2** | SPY | 255 | 246 | 0 | 0 | 50.7 | 1 | 0 |
| **3** | SPY | 189 | 179 | 67 | 65 | 50.8 | 0.744094488 | 0.272357724 |
| **4** | SPY | 178 | 161 | 85 | 75 | 52.3 | 0.703557312 | 0.345528455 |
| **ensemble** | SPY | 208 | 197 | 49 | 47 | 51.09 | 0.815686275 | 0.199186992 |
| **2** | TGT | 179 | 182 | 61 | 79 | 47.9 | 0.69379845 | 0.251028807 |
| **3** | TGT | 222 | 217 | 26 | 35 | 49.2 | 0.86381323 | 0.106995885 |
| **4** | TGT | 127 | 150 | 93 | 129 | 43.69 | 0.49609375 | 0.382716049 |
| **ensemble** | TGT | 199 | 204 | 39 | 59 | 47.32 | 0.771317829 | 0.160493827 |

**Question 4-8.**

As you can see from the consolidated table above, the TPR and other rates related to positive values tend to have the greatest effect on the accuracy. Since the counts for negatives are not very high, it makes sense that positives hold more weight. As TPR increases, accuracy increases. Although TNR doesn’t put in a lot of effect when it is higher, since the TPR is usually high, it brings the accuracy down when TNR is lower.

**Question 5-1.** See below for code-generated plot with x axis as the date from 2021-2022 and y-axis as the amount in dollars.

**Chart

Description automatically generated**

**Question 5-2.**

It is pretty clear to see that all strategies lead to gains in the beginning and a major loss towards the end of the 2 year period. It seems the buy-and-hold strategy is the best of the three, while still ending up losing money, it shows a greater ending point than the rest. Oddly enough, what should have been the most efficient strategy, ensemble, looks like it is the worst strategy with the greatest loss.