Linear Regression

Para 1: First we import all the relevant libraries including matplotlib, numpy, pandas, math, seaborn and time. I defined the input parameters for test_size, cv_size and our stk_path. I expected the code to run properly without any logical errors. Initially, I found out I didnt have some of the libraries like seaborn, time, etc. But after downloading, everything was fine.

Para 2: Second we define a dataframe to get prediction at timestep t using values from t-1, t-2,... t-N. At last, we are computing the function to get mape.

Para 3: Third now we store the data from csv file into our dataframe df. We make some necessary adjustments. And then we explore the first 10 data sets. I expected the code to run properly and give me the necessary close price data from the first ten data points of the csv file. And luckily, it did.

Para 4: Fourth now we plot adjusted close price over time. The plot came out to be expectedly fine.

Para 5: Fifth now we split the dataset into train, validation and test sets in 60%, 20% and 20% respectively. We start by getting the sizes of each of the datasets and then we proceed to split the data into train, validation and test sets. This step also went fine.

Para 6: Sixth now we plot the adjusted close price over time depicting train, validation and test sets with different colours. The graph came out just fine.

Para 7: Seventh now we try to predict validation datapoints. Lastly, we print RMSE, R2 and MAPE for those specific datapoints. And then we show the first five values in our validation dataset. It went as planned and gave all the necessary output.

Para 8: Eighth now we plot RMSE vs N to determine optimum N. We get that N=1 is optimum as the RMSE is least.

Para 9: Ninth now we plot R2 vs N to determine optimum N. Larger the R2 optimum is the N.

Para 10: Tenth we plot MAPE vs N to determine optimum N. Smaller the MAPE optimum is the N.

Para 11: Eleventh we set N_opt = 5.

Para 12: Twelfth now just for checking we plot predictions for a specific day. The graph came out as expected.

Para 13: Thirteenth now we plot predictions on validation dataset. It took some tries but eventually the graph came about after adjusting some of the parameters.

Para 14: Fourteenth now we plot the above plot again but this time zooming into the validation dataset. It went as expected.

Para 15: Fifteenth now we create a final model with test dataset and find out RMSE, R2 and MAPE and bring out the five initial datapoints of test dataset.

Para 16: Sixteenth now we plot the adjusted close vs time using the predictions for Nopt = 5 for our test dataset. The plot came out as expected.

Para 17: Seventeenth now we again plot the above thing while zooming into test dataset. Luckily it went well.

Para 18: Eighteenth now we plot adjusted close vs time with test dataset while using predictions from linear regression. I had to adjust the ylim in order to conceal the plot in the screen without cuts.

Para 19: Lastly, we save our test predicted dataset into a csv file.

Moving Average

Para 1: First we import all the relevant libraries including matplotlib, numpy, pandas, math, seaborn and time. I defined the input parameters for test_size, cv_size and our stk_path. I expected the code to run properly without any logical errors. Initially, I found out I didnt have some of the libraries like seaborn, time, etc. But after downloading, everything was fine.

Para 2: Second we define a dataframe to get prediction at timestep t using values from t-1, t-2,... t-N. At last, we are computing the function to get mape.

Para 3: Third now we store the data from csv file into our dataframe df. We make some necessary adjustments. And then we explore the first 10 data sets. I expected the code to run properly and give me the necessary close price data from the first ten data points of the csv file. And luckily, it did.

Para 4: Fourth now we plot adjusted close price over time. The plot came out to be expectedly fine.

Para 5: Fifth now we split the dataset into train, validation and test sets in 60%, 20% and 20% respectively. We start by getting the sizes of each of the datasets and then we proceed to split the data into train, validation and test sets. This step also went fine.

Para 6: Sixth now we plot the adjusted close price over time depicting train, validation and test sets with different colours. The graph came out just fine.

Para 7: Seventh now we try to predict validation datapoints. Lastly, we print RMSE and MAPE for those specific datapoints. And then we show the first five values in our validation dataset. It went as planned and gave all the necessary output.

Para 8: Eighth now we plot RMSE vs N to determine optimum N. We get that N is optimum as the RMSE is least.

Para 9: Ninth we plot MAPE vs N to determine optimum N. Smaller the MAPE optimum is the N.

Para 10: Tenth we set N_opt = 2.

Para 11: Eleventh now just for checking we plot predictions for a specific day. The graph came out as expected.

Para 12: Twelfth now we plot predictions on validation dataset. It took some tries but eventually the graph came about after adjusting some of the parameters.

Para 13: Thirteenth now we plot the above plot again but this time zooming into the validation dataset. It went as expected.

Para 14: Fourteenth now we create a final model with test dataset and find out RMSE, R2 and MAPE and bring out the five initial datapoints of test dataset.

Para 15: Fifteenth now we plot the adjusted close vs time using the predictions for Nopt = 2 for our test dataset. The plot came out as expected.

Para 16: Sixteenth now we again plot the above thing while zooming into test dataset. Luckily it went well.

Para 17: Seventeenth now we plot adjusted close vs time with test dataset while using predictions from linear regression. I had to adjust the ylim in order to conceal the plot in the screen without cuts.

Para 18: Lastly, we save our test predicted dataset into a csv file.

Last Value

Para 1: First we import all the relevant libraries including matplotlib, numpy, pandas, math, seaborn and time. I defined the input parameters for test_size, cv_size and our stk_path. I expected the code to run properly without any logical errors. Initially, I found out I didnt have some of the libraries like seaborn, time, etc. But after downloading, everything was fine.

Para 2: Second we define a dataframe to get prediction at timestep t using values from t-1, t-2,... t-N. At last, we are computing the function to get mape.

Para 3: Third now we store the data from csv file into our dataframe df. We make some necessary adjustments. And then we explore the first 10 data sets. I expected the code to run properly and give me the necessary close price data from the first ten data points of the csv file. And luckily, it did.

Para 4: Fourth now we plot adjusted close price over time. The plot came out to be expectedly fine.

Para 5: Fifth now we split the dataset into train, validation and test sets in 60%, 20% and 20% respectively. We start by getting the sizes of each of the datasets and then we proceed to split the data into train, validation and test sets. This step also went fine.

Para 6: Sixth now we plot the adjusted close price over time depicting train, validation and test sets with different colours. The graph came out just fine.

Para 7: Seventh now we try to predict validation datapoints. Lastly, we print RMSE and MAPE for those specific datapoints. And then we show the first five values in our validation dataset. It went as planned and gave all the necessary output.

Para 8: Eighth now we plot RMSE vs N to determine optimum N. We get that N is optimum as the RMSE is least.

Para 9: Ninth we plot MAPE vs N to determine optimum N. Smaller the MAPE optimum is the N.

Para 10: Tenth we set N_opt = 2.

Para 11: Eleventh now just for checking we plot predictions for a specific day. The graph came out as expected.

Para 12: Twelfth now we plot predictions on validation dataset. It took some tries but eventually the graph came about after adjusting some of the parameters.

Para 13: Thirteenth now we plot the above plot again but this time zooming into the validation dataset. It went as expected.

Para 14: Fourteenth now we create a final model with test dataset and find out RMSE, R2 and MAPE and bring out the five initial datapoints of test dataset.

Para 15: Fifteenth now we plot the adjusted close vs time using the predictions for Nopt = 2 for our test dataset. The plot came out as expected.

Para 16: Sixteenth now we again plot the above thing while zooming into test dataset. Luckily it went well.

Para 17: Seventeenth now we plot adjusted close vs time with test dataset while using predictions from linear regression. I had to adjust the ylim in order to conceal the plot in the screen without cuts.

Para 18: Lastly, we save our test predicted dataset into a csv file.