FORECASTING MSFT DATA

MANAS BEDMUTHA

Data

- Train Time - 1st Jan 2016 - 1st Jan 2018
- Test Time - 1st Jan 2018 - 1st July 2018
- Features considered Daily values of
  - High
  - Low
  - Closing
  - Opening
  - Adj Closing
- Target value = Adj Close
- This is chosen to account for any corporate actions made during the day
- Data downloaded from Yahoo Finance

Assumptions

- Each price is loosely affected based on its value over a month (taken 20 days; 4 weeks)
- More contribution is due to its fluctuations in the past week (5 days)
- Effect of external factors such as government policies or international effects is uniferrable and ignored
- All data is treated as a regular time series

Approach

- Fluctuations are calculated by first order differences in values of Open, Close, High, Low and Adj Close
- New data frame is collected consisting of moving window averages of above values.
- Window size is 20 for monthly considerations and 5 for weekly considerations of differences and NaN values are made zero
- Intuition is to find daily fluctuation in value to map to target variable

Approach

- These 10 features are fed into a linear regression model (found to give best results among those tried)
- Each such difference is appended to time series based on rolling window averages over last 20 working days
- Output is taken across the features created as above.

Code and Results

- RMS deviation of 27 was found over the test range
- Libraries needed are only sklearn, numpy and matplotlib
- In the attached notebook, inc represents +/- 1 or 0 for increase, decrease or unchanged for dates corresponding to MSFT_test.csv
- Dates have gaps for weekend s which have not been corrected due to shortage of time.