Applied Machine Learning

Security Price Prediction

Application of Regression and Classification Algorithms

Security Price Prediction Problem

- Download a broad market-wide index from a market (e.g., Nifty 50 from India, S&P500 from the US, FTSE100 from the UK, etc.) for 3-5 years (Yahoo Finance or any other public source available).
- Also, download 5-10 securities' data from the same market. These securities must be part of some broad market-wide index from the same market (e.g., Nifty50 or BSE-Sensex from India, S&P500 or NYSE index from the US, etc.)
- Calculate and visually examine the returns of each security individually and also for the market index returns.
- Compute equal-weighted portfolio returns from these 5-10 securities (Assigning equal weights to returns, excluding the market index).

Security Price Prediction Problem: Regression Problem

- Segregate the dataset into training and test dataset. Train dataset with 80% observations and test dataset with 20% observations.
- Train the regression algorithm considering the market return variable as the only feature (independent variable) and the equal-weighted portfolio returns as the dependent variable.
- Evaluate the trained algorithm (using in-sample fitted values) with various goodness-of-fit measures and through visualization.

Security Price Prediction Problem: Regression Problem

- Evaluate the trained algorithm on the out-of-sample dataset for its efficiency in out-of-sample prediction and through visualization.
- Summarize your learnings in the form of a report and also prepare a short presentation for submission.

Security Price Prediction Problem: Classification Problem

- Using an equal-weighted portfolio (5-10 securities) over the sample period, create the dependent variable (or the binary response variable) and name it 'UPDOWN': 1' for positive returns and 'UPDOWN': 0' for negative returns.
- Segregate the dataset into training and test dataset. Train dataset with 80% observations and test dataset with 20% observations.
- Train the linear, logit, and probit algorithms considering the market return variable as the only feature (independent variable) and the UPDOWN variable as the dependent variable.

Security Price Prediction Problem: Classification Problem

- Evaluate and compare the performance of trained linear, logit, and probit algorithms (using in-sample fitted values) with various goodness-of-fit measures, sensitivity, specificity, and accuracy, and through visualization (use threshold value of 0.4, 0.5, 0.6, 0.7 wherever needed).
- Construct the receiver operating characteristic (ROC) curve and compute the area under the curve. Briefly discuss the analysis, learnings, and implications in the report and make a short presentation on it.