CQF Exam Three

Machine Learning

June 2023 Cohort

Instructions: The submitted report must present work and outputs clearly separated by Question. Submit ONLY ONE zip file named LASTNAME.zip that includes pdf file, code, html, data and any other supporting or working files. Python notebook with auxiliary output (data, plots) is not an analytical report: such submission will receive a deduction.

Please do not discuss this assignment in groups or messengers. Raise a support ticket for your queries. Only clarifying questions are allowed.

Introduction: Short-term asset return is a challenging quantity to predict. Efficient markets produce near-Normal daily returns with no significant correlation between r_t, r_{t-1}. This exam is a limited exercise in supervised learning. You are expected to explore multiple features or may use a set of features from Table 1 without an expectation of predictive powers. Original and final selected features should be sufficiently large.

Objective

Your objective is to produce a model to predict positive moves (up trend) using machine learning model as specified in the below section. Your proposed solution should be comprehensive with the detailed feature engineering and model architecture.

- Choose one ticker of your interest from the index, equity, ETF, crypto token, or commodity.

- Predict trend only, for a short-term return (example: daily, 6 hours). Limit prediction to binomial classification: the dependent variable is best labeled [0, 1]. Avoid using [-1, 1] as class labels.

- Analysis should be comprehensive with detailed feature engineering, data pre-processing, model building, and evaluation.

  Devise your own approach on how to categorise extremely small near-zero returns (drop from training sample, group with positive/negative). The threshold will strongly depend on your ticker. *Example:* small positive returns below 0.25% can be labelled as negative.

Table 1: Features List

Feature	Formula	Description
O-C, H-L	Open - Close, High - Low	intraday price range
Sign	sign $[rt = \ln \frac{P_t}{P_{t-1}}]$	sign of return or momentum
Past Returns		lagged returns
Momentum		price change over k period
Moving Average	$\cdots_{t-i}$	simple moving average
Exponential MA	$\cdots + \alpha[P_t - EMA_{t-1}]$	recursive, $\alpha = 2/(N_{obs} + 1)$

Number of features to include is a design choice. There is no one recommended set of features for all assets. Length of dataset is another design choice. If predicting short-term return sign (for daily move), then training and testing over up to 5-year period should be sufficient. Making sense of instructions below is part of the task. The tutor will not assist in designing your computational implementation.

Questions

1. What are voting classifiers in ensemble learning?

2. Explain the role of the regularization parameter C in a Support Vector Machine (SVM) model. How does varying C affect the model's bias and variance trade-off?

3. Follow the 7-steps to model building for your selected ticker,

  (a) produce a model to predict positive moves (up trend) using Random Forest Classifier.

  (b) tune hyperparameters for the estimator and present the best model.

  (c) investigate the prediction quality using area under ROC curve, confusion matrix and classification report.

Note: Choice and number of hyperparameters to be optimized for the best model are design choices. Use of experiment tracking tools like MLFlow is allowed [refer to Advanced Machine Learning Workshop - II for sample implementation].

$* * *$