

# CQF Exam Three

## Machine Learning

June 2022 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. Address all portal and upload queries to CQFProgram@fitchlearning.com and clarifying only questions to Kannan.Singaravelu@fitchlearning.com.

**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: use a set of features from Table 1 without an expectation of predictive powers.

- Choose one ticker of your interest from: equity, ETF, crypto token, or commodity. Do not choose: FX tickers (GBPUSD), equities with market cap over 100 bln. USD.
- Predict trend only, for a short-term return (example: daily, 6 hours). Limit prediction to binomial classification: dependent variable is best labelled  $[0, 1]$ . Avoid using  $[-1, 1]$  as class labels.

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
<i>O-C, H-L</i>	Open - Close, High - Low	intraday price range
<i>Sign</i>	$\text{sign} [rt = \ln \frac{P_t}{P_{t-1}}]$	sign of return or momentum
<i>Past Returns</i>	$r_{t-1}, r_{t-2}, \dots$	lagged returns
<i>Momentum</i>	$P_t - P_{t-k}$	price change over k period
<i>Moving Average</i>	$SMA_i = \frac{1}{n} \sum_{i=0}^{n-1} P_{t-i}$	simple moving average
<i>Exponential MA</i>	$EMA_i = EMA_{t-1} + \alpha [P_t - EMA_{t-1}]$	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.

## A. Maths

1. Consider  $MSE(\hat{\beta})$  wrt to the true value  $\beta$  in context of regression methods,

$$E[(\hat{\beta} - \beta)^2] = Var[\hat{\beta}] + (E[\hat{\beta}] - \beta)^2$$

Answer below with Yes/No and one sentence of explanation referring to maths.

- (a) can there exist an estimator with the smaller MSE than minimal least squares?
- (b) for a prediction, does the MSE measure an irreducible error or model error?
2. Gaussian RBF kernel is given as  $k(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma}\right)$ . Suppose we have three points,  $z_1, z_2$  and  $x$ ; where  $z_1$  is geometrically very close to  $x$ , and  $z_2$  is geometrically far away from  $x$ . What is the value of  $k(z_1, x)$  and  $k(z_2, x)$ ? Choose the correct answer below and explain it with reasoning.
- (a)  $k(z_1, x)$  will be close to 1 and  $k(z_2, x)$  will be close to 0.
- (b)  $k(z_1, x)$  will be close to 0 and  $k(z_2, x)$  will be close to 1.
3. Mathematically specify three types of Loss Function for decision trees.

## B. Models

4. Follow the 7-steps to model building for your selected ticker,
- (a) produce a model to predict positive moves (up trend) using Support Vector Machines.
- (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.

**Add WeChat powcoder**

Note: Choice of kernels 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].

\* \* \*