

① You are about to start the Certificate in Quantitative Finance (CQF) exam. By accessing the exam materials, you acknowledge and agree to the CQF Terms of Use (<https://www.fitch.group/terms-of-use/Links to an external site.>) and the following terms:

- Fitch Learning Limited (“Fitch Learning”) holds ownership or licenses for all intellectual property rights related to the www.cqf.com website and any content available on the website and the exam, including text, graphics, software, photographs, images, videos, sound, trademarks, and logos (the “CQF Materials”).
- By clicking "Start Assignment," you consent not to infringe upon Fitch Learning’s intellectual property rights. You agree not to copy, modify, reproduce, publish, sell, upload, broadcast, post, transmit, license, distribute, adapt, merge, translate, disassemble, decompile, or reverse engineer any part of the CQF Materials without Fitch Learning’s written consent.
- Fitch Learning reserves the right to take action against any unauthorized use of the CQF Materials in the event of a breach of these terms. This may include, but is not limited to: (i) disabling and terminating your account; (ii) requesting you to remove materials that breach these terms; and/or (iii) commencing legal proceedings to protect Fitch Learning’s or its licensors’ intellectual property rights.

CQF Exam Three

Machine Learning

June 2025 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 returns are challenging 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 of your choice, with both the original and final selected features being sufficiently numerous.

Objective

Your objective is to develop a model to predict positive market moves (uptrend) using machine learning techniques as outlined in the section below. Your proposed solution should be comprehensive, including detailed feature engineering and model architecture.

- Choose one ticker of your choice from the index, equity, ETF, crypto token, or commodity.
- Predict the trend for short-term returns using binomial classification. The dependent variable should be labeled as $[0, 1]$, not $[-1, 1]$.
- The analysis should be comprehensive, including detailed data preprocessing, feature engineering, model building, tuning, and evaluation.

Devise your own approach for categorizing extremely small near-zero returns (e.g., drop from the training sample or group with positive/negative returns). The threshold will depend on your chosen ticker. *Example:* small positive returns below 0.25% can be labeled as negative.

The number of features to include is a design choice, and there is no universally recommended set of features for all assets. The length of the dataset is also a design choice. For predicting short-term returns (e.g., daily moves), training and testing over a period of up to 5 years should be sufficient. Interpreting the instructions below is part of the task; the tutor will not assist in designing your computational implementation.

A. Maths [20 marks]

1. 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.

2. What are voting classifiers in ensemble learning?

B. Feature Selection Using the Funnelling Approach [20 marks]

3. Perform feature selection for a machine learning model using a multi-step process by combining techniques from filter, wrapper, and embedded methods.

(a) Explain the feature selection process using the three categories of feature selection methods, step by step.

(b) Justify the selection of features retained at each step.

(c) Provide the final list of selected features.

C. Model Building, Tuning and Evaluation [60 marks]

4. Predicting Positive Market Moves Using Support Vector Support Vector Machine (SVM),

(a) Build a model to predict positive market moves (uptrend) using the feature subset derived above.

(b) Tune the hyperparameters of the estimator to obtain an optimal model.

(c) Evaluate the model's prediction quality using the area under the receiver operating characteristic (ROC) curve, confusion matrix, and classification report.

Note: The choice of kernel and the selection of hyperparameters to optimize are critical design decisions in developing an effective model. Submitting Python code alone, without clear explanations or context, will not be accepted. The report must present a detailed study, including methodology, analysis, and a well-reasoned conclusion. As an optional enhancement, you may consider backtesting the predicted signals within a trading strategy to assess their practical impact.

* * *