Final Exam

Machine Learning for Finance (2023-2024)

Examination Details and Instructions

Instructions:

- **Duration**: This exam is 2 hours long.
- **No Devices**: Use of computers, smartphones, or any internet-enabled devices is prohibited.
- **No Notes**: Use of personal notes is forbidden.
- **Evaluation of Research Traces**: Evidence of research will be taken into account in the grading.
- **Neural Network Architecture**: For the architectures requested in Questions 8 and 12, you may illustrate each with a diagram.

Description: The exam is graded on a 100 point scale and is divided into three independent parts. Below is the mark distribution for each question:

Problem	Question	Number of Marks
Problem A	Question 1	5
	Question 2	15
	Question 3	15
	Question 4	5
Problem B	Question 5	15
	Question 6	5
Problem C	Question 7	4
	Question 8	10
	Question 9	3
	Question 10	5
	Question 11	6
	Question 12	12

Please read the questions carefully and do your best. Good luck!

1 Problem A: Fama-French Three-Factor Model (40 marks)

This exercise will guide you through the application of the Fama-French three-factor model to Apple Inc. (AAPL) stock returns. For each time step $t \in [0, T]$, the return r_t of the AAPL stock can be expressed in the Fama French Three Factor model as:

$$r_t - r_{ft} = \alpha_{AAPL} + \beta_M(r_{Mt} - r_{ft}) + \beta_{SMB}SMB_t + \beta_{HML}HML_t + \epsilon_t \tag{1}$$

where r_{ft} is the risk-free rate at time t, r_{Mt} is the market return, SMB_t is the small minus big size premium, and HML_t is the high minus low value premium and ϵ_t an idiosyncratic risk of AAPL at time t that follow a $\mathcal{N}(0, \sigma^2)$ independent of time t.

Question 1 — 5 Marks

Give the name of the machine learning model described in Equation 1.

Question 2 — 15 Marks

Reformulate the Fama-French Model in matrix notation for the period from $t = 0$ to T ($T+1$ steps) as indicated in the following Equation 2.

$$\hat{R}_T = X_T \hat{\beta} + \tilde{\epsilon}_T \tag{2}$$

Clarify how these matrices interact within the model, ensuring the dimensions align correctly for matrix operations from the beginning to the end of the specified time interval.

Please include the following details in your response:

- Describe the matrix X_T and explain what each column represents.
- Define the shapes of $\hat{R}_T$, $\hat{\beta}$, and $\tilde{\epsilon}_T$ in the context of this model.
- Explain how these matrices relate to the original variables in the Fama-French model.

Question 3 — 15 Marks

Prove the equivalence between maximum likelihood estimation (MLE) and ordinary least squares (OLS) for estimating the parameter $\hat{\beta}$ in a linear regression model.

Consider the OLS problem defined as:

$$\min_{\hat{\beta}} \| R_T - X_T \hat{\beta} \|^2 \tag{3}$$

You can follow these steps to establish the equivalence:

- Assume that the residuals $\epsilon_T = R_T - X_T \hat{\beta}$ are normally distributed with mean zero and variance $\sigma^2 I$, where I is the identity matrix. Write the likelihood function for this setup.
- Derive the log-likelihood function from the likelihood function. Simplify the expression.
- Perform the differentiation of the log-likelihood function with respect to β and set the derivative to zero to find the maximum likelihood estimate of β.
- Conclude that this derivation leads to the same solution as the minimization problem in Equation 3.

Question 4 **5 Marks**

In our model the expected prediction error at a point x can be decomposed into bias, variance, and noise according to the following equation:

$$\text{MSE} = \text{Bias}^2(\hat{r}) + \text{Var}(\hat{r}) + \sigma^2 \tag{4}$$

where $\hat{r}$ is the prediction made by the model at x, $\text{Bias}(\hat{r})$ is the difference between the expected prediction and the true value, $\text{Var}(\hat{r})$ is the variance of the prediction at x, and σ^2 represents irreducible error or noise.

- Explain the concepts of bias and variance in this setting and discuss their trade-off.
- Suggest strategies to minimize bias and strategies to reduce variance in model predictions.

2 Problem B: Decision Tree (20 marks)

Imagine you are working with a dataset intended for a binary classification task. The dataset contains the following features: Age, Salary, and Education Level. Your target variable is *Purchased*, which is binary (1 for yes, 0 for no).

The training data snapshot is as follows:

Age	Salary	Education Level	Purchased
25	$50,000	High School	0
40	$65,000	Bachelor's	1
30	$80,000	Bachelor's	0
35	$95,000	Master's	1
45	$110,000	Doctorate	1

Question 5 **15 Marks**

Divide the dataset into two groups based on the Salary being either less than or equal to $90,000, and more than $90,000. Calculate the Gini index for each group and the overall Gini index for the split.

Hint: To calculate the Gini index for a group:

$$\text{Gini} = 1 - (p_1^2 + p_0^2)$$

where p_1 and p_0 are the proportions of class 1 and class 0 in the group, respectively.

To compute the overall Gini index for the split:

$$\text{Gini}_{\text{split}} = \left(\frac{n_A}{n}\right)\text{Gini}_A + \left(\frac{n_B}{n}\right)\text{Gini}_B$$

where n_A and n_B are the number of samples in each group, and n is the total number of samples.

Question 6 **5 Marks**

Consider how different educational levels might affect the likelihood of purchasing, and propose a split. Discuss why this split might make sense from an information gain perspective, focusing on how it could help segregate the data into more homogeneous subsets without using mathematical terms.

3 Problem C: Sentiment Analysis (40 marks)

Consider a dataset for sentiment analysis related to stock price prediction. The dataset comprises N sentences of **varying lengths**, each annotated with one of three sentiments: *Good*, *Neutral* or *Bad*. These sentences are extracted from sources such as news articles, social media posts, or financial reports.

The dataset used for this exercise is as follows:

Index	Sentence	Sentiment
1	Company profits soar.	*Good*
2	Market stability remains uncertain.	*Neutral*
⋮	⋮	⋮
N	Legal issues impact stock prices.	*Bad*

3.1 FeedForward Neural Network

To effectively utilize a Deep Neural Network with 3 dense layers, including the last one, for our classification task, we need to preprocess the data appropriately.

Question 7 **4 Marks**

Describe four preprocessing steps necessary to transform our sentences into fixed-length vectors (of length V).

Question 8 **10 Marks**

For this task, your goal is to design a neural network that classifies sentiments into three categories: *Good*, *Neutral* and *Bad*. Provide a detailed description of your design by addressing the following points:

Neural Network Architecture

- **Input Layer:**
  - Specify the expected dimension V of input vectors, explaining how V is determined based on preprocessing steps.
- **Dense Layers:**

  - Detail the number of units and the type of activation functions used in these layers.
  - Explain why you chose these specific configurations.

- **Output Layer:**
  - State the number of units and describe the activation function used. Explain how this configuration aligns with the task of classifying sentiments into three categories.

Activation Functions

- Clarify the activation functions used in different parts of the network.
- Discuss the rationale behind choosing each activation function, particularly focusing on the output layer.

Data Tensor Shape

- Describe the shape of the data tensor that feeds into the network.
- Explain the significance of each dimension in the tensor.

Question 9 — 3 Marks

Specify the suitable loss function necessary for our neural network model.

Question 10 — 5 Marks

After the first training, we obtain the Figure 1.

Explain what happened and give two possible reasons for a such outcome.

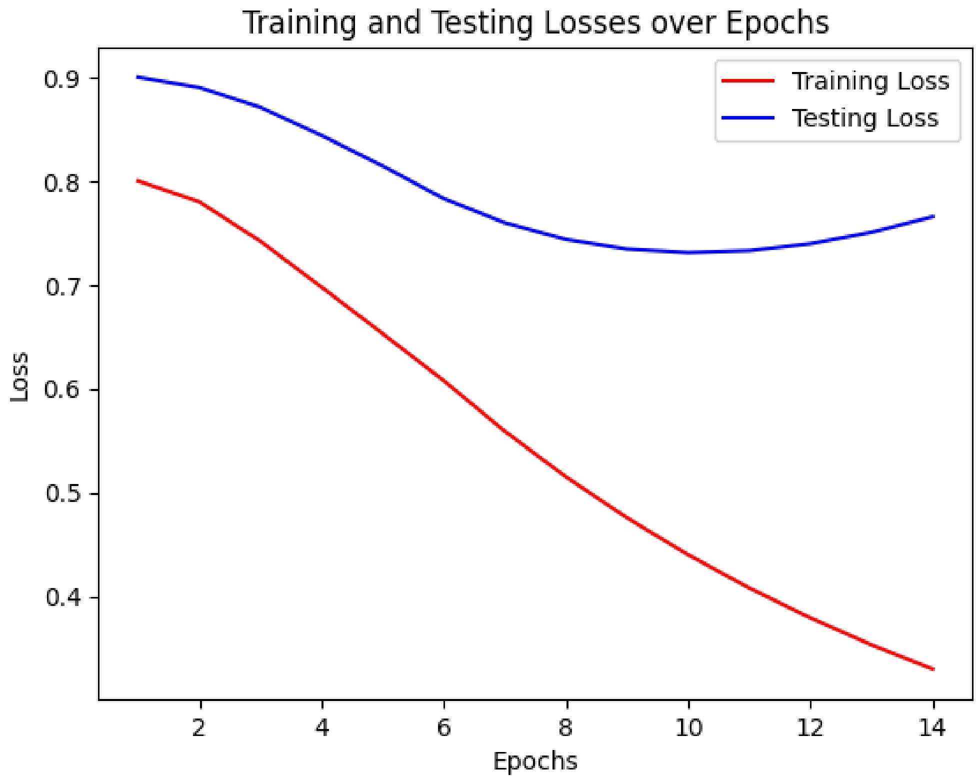

Figure 1: Training and Testing Losses over Epochs

3.2 Sequential Neural Network

We aim to introduce Long Short-Term Memory (LSTM) models instead of using FeedForward network for conducting sentiment analysis.

Question 11 **6 Marks**

Explain the functionality of the gates within an LSTM architecture and how they mitigate the vanishing gradient problem inherent in traditional recurrent neural networks.

Question 12 **12 Marks**

Assume we want to develop a multi-layer LSTM architecture where each of the d LSTM layers outputs a vector of dimension d_p. Provide a comprehensive description of the proposed classification model architecture by addressing the following points:

- **Input Data Shape**:
  - Describe the initial shape of the input data. What does each dimension represent in the context of the problem?

- **Embedding Layer**:
  - Explain how the input shape is transformed by the embedding layer. What is the new shape and what does each dimension represent?
  - Detail the role of the embedding matrix (e.g., Word2vec) and its configuration (fixed or trainable).
- **LSTM Layers**:
  - Describe the structure and function of the LSTM layers in the network.
  - For each LSTM layer, explain how it transforms the shape of the data. What are the dimensions after each layer, and why is this transformation important for the model?
- **Final LSTM Layer**:
  - Discuss the output of the final LSTM layer. How does it differ from the previous LSTM layers in terms of data transformation and output shape?
- **Dense Output Layer**:
  - Specify the configuration of the output layer (number of units and activation function).
  - Explain how this layer transforms the output of the final LSTM layer into the final classification output. What is the final shape of the output tensor?
- **Summary of Shape Transformations**:
  - Provide a concise summary of the transformations from input to output. Include a brief description of the shape at each stage and the significance of these transformations for achieving the model's objective.