Strathmore
UNIVERSITY

STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES
MASTER OF SCIENCE IN DATA SCIENCE & ANALYTICS
END OF SEMESTER EXAMINATION
DSA 8401 APPLIED MACHINE LEARNING IN DATA SCIENCE

Date: 9th December 2024 Time: 3 Hours

Instructions

1. This examination consists of **FIVE** questions.
2. Answer **Question ONE (COMPULSORY)** and any other **TWO** questions.

Question ONE (COMPULSORY) [30 marks]

a)
- i. Differentiate between a model parameter and a learning algorithm's hyperparameter. Provide one example of each. **(2 marks)**
- ii. Explain what model-based learning algorithms search for during training.? **(2 marks)**
- iii. What is the most common strategy used by these algorithms to succeed in finding a good solution? **(2 marks)**

b)
- i. Which Gradient Descent algorithm (among those we discussed) will reach the vicinity of the optimal solution the fastest? (**2 marks)**
- ii. Which will actually converge? **(2 marks)**
- iii. How can you make the others converge as well? **(2 marks)**

c)
- i. Identify two machine learning algorithms that may perform poorly if features have different scales. Explain why? **(2 marks)**
- ii. Discuss two methods to address the issue of different feature scales and explain how they work. **(2 marks).**

d)
- i. Describe the fundamental idea behind Support Vector Machines**. (2 marks)**
- ii. Define a support vector and explain its role in SVM **(2 mark)**

e)
- i. If a Decision Tree is overfitting the training data, list two hyperparameters you can adjust and describe their effect. **(2 marks).**
- ii. If the model is underfitting, suggest two potential changes to the tree and explain why these might help. **(2 marks)**

f)
- i. What are the two motivations for reducing a dataset's dimensionality? **(2 marks)**
- ii. List two potential drawbacks of dimensionality reduction **(2 marks)**
- iii. Explain why dimensionality reduction may or may not be suitable when working with neural networks. **(2 marks)**

Question TWO [15 Marks]

a)

i. Define the Silhouette coefficient and explain what it measures **(3 marks).**

ii. How can we use it to select the right number of clusters when using K-Means? (**2 marks).**

b) Suppose you have an MLP composed of one input layer with 10 passthrough neurons, followed by one hidden layer with 50 artificial neurons, and finally one output layer with 3 artificial neurons. All artificial neurons use the ReLU activation function:

i. What is the shape of the input matrix **X**? **(2 marks)**

ii. What are the shapes of the hidden layer's weight vector **W**_*h* and its bias vector **b**_*h*? **(3 marks)**

iii. What are the shapes of the output layer's weight vector **W**_*o* and its bias vector **b**_*o* (**3 marks)**

iv. What is the shape of the network's output matrix **Y**? **(1 mark)**

v. Write the equation that computes the network's output matrix **Y** as a function of **X**, **W**_*h*, **b**_*h*, **W**_*o*, and **b**_*o* . **(1 mark)**

Question THREE [15 Marks]

a)

i. Briefly describe how backpropagation works? **(3 marks)**

ii. How does it ensure that weights are updated correctly? **(2 marks)**

b) Name three advantages of the SELU activation function over ReLU. **(3 marks**)

c) Consider a CNN composed of three convolutional layers, the first one with 5x5 kernels and the other two with 3 × 3 kernels, a stride of 2, and "same" padding. The lowest layer outputs 50 feature maps, the middle one outputs 100, and the top one outputs 200. The input images are RGB images of 200 × 300 pixels. What is the total number of parameters in CNN? **(7 marks)**

Question FOUR [15 Marks]

a) Sketch the LSTM cell's architecture and describe the function of the elements. **(6 marks)**

b) Suppose you want to train a classifier, and you have plenty of unlabeled training data but only a few thousand labeled instances. How can autoencoders help? How would you proceed? **(5 marks)**

c) Explain how Dropout helps in training neural networks and discuss its limitations. **(4 marks)**

Question FIVE [15 Marks]

a) Explain the difference between bias and variance in machine learning. How do they contribute to the overall error of a model? **(5 marks)**

b)

i. What is the curse of dimensionality, and how does it affect model performance? **(4 marks)**

ii. Suggest two techniques to mitigate its effects. **(3 marks)**

iii. Compare and contrast bagging, boosting, and stacking in ensemble methods. (**3 marks)**