



(Knowledge for Development)

**KIBABII UNIVERSITY
(KIBU)**

**UNIVERSITY EXAMINATIONS
2022/2023 ACADEMIC YEAR**

**END OF SEMESTER EXAMINATIONS
YEAR 1 SEMESTER II**

**FOR THE DEGREE OF
MASTER OF SCIENCE IN COMPUTER SCIENCE**

COURSE CODE : MCS 8013

COURSE TITLE : ADVANCED MACHINE LEARNING

DATE: 10/02/2023

TIME: 9.00 AM -12.00 A.M

INSTRUCTIONS

SECTION A IS COMPULSORY.

**ANSWER ANY 2 QUESTIONS FROM SECTION B. EACH QUESTION IN
THIS SECTION CONTAINS 20 MARKS.**

SECTION A (COMPULSORY QUESTION)

QUESTION ONE [20 MARKS]

- 1) Differentiate between Artificial Intelligence, Machine Learning and Deep Learning. (2 Marks)

- 2) Define the following terminologies in relation to machine learning: (6 Marks)
 - a) Generalization Error
 - b) Regularization
 - c) Data augmentation

- 3) Machine learning is driving innovation in many fields, and every day we're seeing new interesting use cases emerge. Explain at least **two** use cases of machine learning (2 Marks)

- 4) Explain at least **two** features of machine learning (2 Marks)

- 5) What is Hyperparameter Optimization? Name at least **three** hyperparameter tuning methods. (3 Marks)

- 6) Name at least **four** most popular open-source libraries for machine learning. (2 Marks)

- 7) Give at least **one** reasons as to why Python is one of the most popular programming languages for working with machine learning and state the functions of *Numpy* and *Pandas* Python libraries when performing data preprocessing using Python. (3 Marks)

SECTION B

ANSWER ANY 2 QUESTIONS FROM THIS SECTION

QUESTION TWO [20 MARKS]

- 1) At a broad level, machine learning can be classified into three types/categories based on how learning is received or how feedback on the learning is given to the system developed. Name these categories and give a brief explanation of each. **(9 Marks)**

- 2) Consider a machine learning model trained till convergence on some training data D^{train} , and tested on separate test set D^{test} . You look at the test error, and find that it is very high. You then compute the train error and find that it is close to 0.
 - a) What could be the possible course of this occurrence? **(1 1/2 Marks)**
 - b) Name and briefly explain at least three approaches that can be used to reduce such occurrence. **(9 Marks)**
 - c) If you plot the train and test errors as a function of the model complexity, How is your plot expected to look like? **(2 Marks)**

QUESTION THREE [20 MARKS]

- 1) Why is data preprocessing crucial step while creating a machine learning model? Outline the steps involved in data preprocessing. **(5 Marks)**

- 2) Explain the following terminologies as used in regression. **(6 Marks)**
 - a) Cost function
 - b) Gradient Descent
 - c) Polynomial Regression

- 3) Briefly explain the following classification algorithms **(6 Marks)**
 - a) Logistic Regression
 - b) Support Vector Machines
 - c) K-Nearest Neighbours

- 4) What is feature selection? Give one benefit of feature selection and give one techniques that can be used in feature selection. **(3 Marks)**

QUESTION FOUR [20 MARKS]

- 1) A wide range of supervised learning algorithms are available, each with its strengths and weaknesses. In order to choose the best algorithm for your model, there are a number of major factors to consider. One of the issue to consider is Bias-variance tradeoff.
 - a) Discuss the concept of Bias-Variance tradeoff. **(8 Marks)**
 - b) Give at least **three** examples of high-bias machine learning algorithms and at least **three** examples of low-bias machine learning algorithms. **(3 Marks)**
 - c) Besides Bias-Variance tradeoff, explain three other factors to consider when choosing an algorithm for your model. **(9 Marks)**

QUESTION FIVE [20 MARKS]

- 1) Differentiate between the following:
 - a) Irreducible errors and reducible errors **(2 Marks)**
 - b) Overfitting and underfitting **(3 Marks)**
 - c) Empirical risk minimization and Structural risk minimization **(3 Marks)**
- 2) Cross-validation is one technique used to prevent overfitting. Explain how cross-validation works and explain at least three methods used for cross-validation. **(12 Marks)**