# EMATM0044 Advanced Topics in Machine Learning Test

This test consists of five questions, each worth ten marks. The final version of this exam will be conducted online. This document is a draft.

## Q1

1. Explain how the Naive Bayes classifier works for text classification. [2 marks]
2. Discuss three assumptions made by the Naive Bayes classifier. [3 marks]
3. Given the following table of email messages, each characterized by words: "free" and "money", and classified as spam or not spam:

| ID | Class | Contains 'Free' | Contains 'Money' |
|---|---|---|---|
| 1 | Spam | Yes | No |
| 2 | Spam | Yes | Yes |
| 3 | Spam | No | Yes |
| 4 | Not Spam | No | No |
| 5 | Not Spam | No | Yes |
| 6 | Not Spam | Yes | No |
| 7 | Unknown | Yes | Yes |

Using Naive Bayes, classify the email at the 7th row as spam or not spam. Explain your process and assumptions. [5 marks]

## Q2

1. Explain what is meant by the terms precision and recall in the context of a classification model. [2 marks]
2. How does the F1 Score integrate precision and recall, and why might it be a more useful measure than either precision or recall alone? [3 marks]
3. Given the following confusion matrix for a prediction model on patient diagnoses:

| | | Predicted | |
|---|---|---|---|
| | | Positive | Negative |
| Actual | Positive | 30 | 5 |
| | Negative | 10 | 55 |

Calculate the precision, recall, and F1 score for the model. Show your work. [3 marks]

4. Discuss the potential consequences of having low precision or low recall in medical diagnostic tests. [2 marks]

## Q3

1. What is logistic regression, and when is it appropriate to use it? [2 marks]
2. Discuss the impact of the logistic function in logistic regression and its importance for modeling probabilities. [2 marks]
3. Consider the following dataset of exam scores and whether students passed or failed:

| Score | Passed |
|---|---|
| 50 | No |
| 55 | No |
| 65 | Yes |
| 70 | Yes |
| 75 | Yes |
| 80 | No |
| 85 | Yes |

   Using logistic regression, explain how you would approach modeling this data to predict whether a new student with a score of 68 would pass or fail. [3 marks]
4. How does regularization affect the training of logistic regression models? Provide an example of a regularization method and explain its effect. [3 marks]

## Q4

Discuss the ethical implications of AI in automated decision systems, specifically addressing:

1. The importance of AI ethics in the context of automated decision-making systems. [3 marks]
2. Identify one potential ethical issue related to AI decision-making in the criminal justice system. [4 marks]
3. Suggest strategies to address the identified ethical issue, focusing on fairness and accountability. [3 marks]

## Q5

1. Describe the concept of heuristic search in artificial intelligence. [3 marks]
2. Compare and contrast the greedy best-first search algorithm with the A* search algorithm. [3 marks]
3. Discuss scenarios where greedy best-first search may lead to suboptimal solutions. [2 marks]

4. If using greedy best-first search to navigate a robot through an obstacle course, what kind of heuristic might be effective? Assume the robot has to avoid dynamic obstacles. [2 marks]