Project-Based Assessment: Applying Fundamental Mathematics in AI Problem Solving

Assignment hand out: 18th September 2023
Hand-in: **Full time students**: 12th October 2023 at 5pm
Part time students (PGCert): 19th October 2023 at 5pm
You will receive feedback within **10 working days** of submission of assessment.

Contribution towards module: 100% (and scored out of 100 marks). This is an individual assessment. When submitting your assignment, you are agreeing to the following statement:

I certify that the submission is my own work, all sources are correctly attributed, and the contribution of any AI technologies is fully acknowledged.

The university policy on plagiarism is available at:
https://www.qub.ac.uk/directorates/sgc/learning/LearningResources/Plagiarism/

Keep a copy of all submitted coursework – i.e., your computer files. **Remember plagiarism is a serious offense** – ensure that you use your own words even when referencing a source. Also reference libraries, tutorials, or code that you used. This assignment is to be completed individually.

Learning outcomes: Demonstrate knowledge and critical understanding of topics from linear algebra, calculus, probability, statistics, and optimisation that are required to apply in AI. Relate the mathematics topics to AI.

Assignment Brief:

You will work on a comprehensive project that involves data analysis, model development, and optimisation on an openly available dataset - German Credit Data from UCI Repository. The data is available at http://archive.ics.uci.edu/dataset/144/statlog+german+credit+data. This data was donated by Hans Hofmann of the University of Hamburg. The dataset contains information on loans obtained from a credit agency in Germany. The primary motivation behind this project is to find the factors associated with an elevated risk of loan default. You are required to write a report detailing your methodology and analysis of this open dataset utilising various mathematical concepts covered in this module and showing results and data visualisations.

This coursework aims to assess your ability to apply foundational mathematical concepts in linear algebra, calculus, probability, statistics, and optimisation to solve AI-related problems. Your task is to develop an AI solution that showcases your proficiency in integrating mathematical techniques. Each week you should be able to incorporate additional components into your report in line with your practical work, lectures and directed reading. As the weeks progress, retain the analysis you have undertaken in earlier weeks, so the report and Python script (or Jupyter Notebook) is your cumulative work.

Project Phases:

Phase 1: Problem Definition and Data Preparation

- Describe the selected dataset and its attributes.
- Identify the specific issue or task you aim to solve using AI techniques. For example, identify the key factors contributing to a higher risk of loan default.
- Preprocess the data including the handling of missing values, encoding categorical variables, and necessary transformations to ensure data quality and suitability for subsequent analyses.

Phase 2: Exploratory Data Analysis

- Perform exploratory analysis on the selected dataset, examining the distribution, summary statistics, and relationships between variables.
- Apply probabilities and statistical analysis techniques, such as hypothesis testing and correlation analysis, to explore relationships and draw meaningful insights.
- Enhance your analysis through the incorporation of data visualisation techniques, using graphs and plots to explore patterns and trends within the data.

Phase 3: Mathematical Foundation Applications

Phase 3.1: Singular Value Decomposition (SVD) or Principal Component Analysis (PCA) for feature reduction

- Apply SVD or PCA to the pre-processed dataset.
- Analyse the principal components obtained and interpret their significance in relation to target prediction.

Phase 3.2: Model optimisation

- Formulate an optimisation challenge aligned with your AI objective (e.g., hyperparameter refinement).
- Employ appropriate optimisation methods (e.g., gradient descent) to identify optimal parameter values.

Phase 3.3: Markov Decision Process for decision optimisation

- Define a simplified Markov Decision Process (MDP) to simulate the process of enhancing decision-making in the context of loan defaults.
- Identify states, actions, rewards, and transition probabilities.
- Implement appropriate learning strategy to optimise decisions for enhancing the decision process.

Phase 4: Discussion and interpretation

- Analyse the results to draw insights.
- Assess the impact of each technique on the final AI solution's performance and decisions.

Project Deliverables:

1. Report

You are required to submit a project report not exceeding **3000 words**. The project requires you to create a comprehensive report that covers all aspects of the project. This report should include various sections including **1) Introduction, 2) Methodology, 3) Results, 4) Discussions and 5) Conclusion & Word Count.** Specifically, elaborate on the outcomes and insights obtained from each Mathematical Foundation Application. You should provide your rationale for selecting the appropriate mathematical techniques. To enhance the report's clarity,

incorporate visualisations, tables, and code snippets that effectively support the analysis and findings you present. Throughout the report, please ensure to provide appropriate references and sources to support your methodology, analysis, and findings. Proper citation of relevant literature and online sources is essential to demonstrate the credibility and validity of your work. Word count excludes Figures, Tables, Titles, References, Code and Cover page.

A margin of +10% of the size limit will be allowed before the penalties for exceeding the word count apply.

Penalties for exceeding the word count follow the University Guidance and are as follows:

- +10% - no penalty
- +>10% - 20% - 10% penalty
- +>20% - 30% - 20% penalty
- +>30% - 40% - 30% penalty
- +>40% - 50% - 40% penalty
- +>50% - maximum mark of 50% for the assignment.
- All deductions won't lower the assignment grade below the passing threshold.

2. Video Presentation

You are also required to submit a **5-minute** video presentation that summarises the crucial aspects and findings of your project. In the video, ensure you provide a clear and thorough explanation of the dataset used, elaborate on the analysis techniques you employed, and present the significant outcomes you achieved. Make sure to highlight the valuable insights that emerged from the utilisation of various mathematical concepts. Your video presentation should effectively communicate the essence of your project's journey and findings.

Submission

The report will be electronically submitted online using Canvas. Your submission should consist of a .zip file containing the report along with all other relevant files. Python code should be well-documented, structured, and include appropriate comments and must be copied directly from your Python script file or Jupyter Notebook file as an appendix of the report for plagiarism check. In addition, please submit the Python script file or Jupyter Notebook file separately as part of your submission. You will be given an individual mark based on the assessment criteria. After submission, please double-check that all submitted work has been uploaded correctly on Canvas.

Standard university penalties apply for late submission -- penalised at the rate of 5% of the total mark awarded for that component for each calendar day late up to a maximum of five calendar days, after which a mark of zero shall be awarded, i.e., up to one calendar day is -5%; up to two calendar days is -10%; up to three calendar days is - 15%, etc. This excludes University closure days.

https://www.qub.ac.uk/directorates/AcademicStudentAffairs/AcademicAffairs/GeneralRegulations/StudyRegulations/StudyRegulationsforPostgraduateTaughtProgrammes/

Assessment Criteria	Marks
Problem Definition and Data Preparation **Outstanding (9-10 marks):** Clear problem definition and concise dataset overview. Thorough data preprocessing: missing values and categorical variables handled effectively.	**10 Marks**

Good (7-8 marks): Well-defined problem and clear dataset description. Effective data preprocessing; minor improvement possible. **Acceptable (5-6 marks):** Acceptable problem definition and dataset summary. Basic data preprocessing with some gaps. **Poor (3-4 marks):** Problem description lacks clarity or dataset overview. Incomplete data preprocessing: key steps omitted. **Very poor (0-2 marks):** Problem statement or dataset description is missing. Very poor data preprocessing; essential steps missing.	
Exploratory Data Analysis **Outstanding (17-20 marks):** Comprehensive data exploration with insightful visualisations. Proficient application of probabilities, hypothesis testing, and correlation analysis. **Good (13-16 marks):** Sound data exploration with clear visuals. Effective use of statistical techniques. **Acceptable (9-12 marks):** Acceptable data exploration with basic visuals. Adequate use of some statistical techniques. **Poor (5-8 marks):** Limited data exploration; visuals lack depth. Limited application of statistical techniques. **Very poor (0-4 marks):** Inadequate data exploration; visuals missing. Inadequate or no application of techniques.	**20 Marks**
SVD or PCA for feature reduction **Outstanding (17-20 marks):** Proficient SVD or PCA application and meaningful component interpretation. **Good (13-16 marks):** Effective SVD or PCA usage and clear component interpretation. **Acceptable (9-12 marks):** Acceptable SVD or PCA application and some component interpretation. **Poor (5-8 marks):** Limited SVD or PCA implementation and component interpretation. **Very poor (0-4 marks):** Inadequate SVD or PCA implementation or component interpretation.	**20 Marks**
Model optimisation **Outstanding (17-20 marks):** Precise formulation of AI-relevant optimisation problem. Proficient optimisation method application with optimal parameters identified. **Good (13-16 marks):** Clear problem formulation with some AI relevance. Effective optimisation technique application with parameter refinement. **Acceptable (9-12 marks):** Adequate optimisation problem definition. Adequate technique application with room for improvement. **Poor (5-8 marks):** Problem formulation lacks clarity or relevance. Limited optimisation technique usage and parameter refinement needed. **Very poor (0-4 marks):** Problem formulation is unclear or missing. Inadequate or no optimisation technique application.	**20 Marks**
Markov Decision Process for decision optimisation **Outstanding (9-10 marks):** Comprehensive MDP formulation with clear components. Proficient state-action dynamics definition; effective learning strategy. **Good (7-8 marks):** Clear understanding of MDP concept; solid formulation. Clear state-action dynamics and reasonable learning strategy. **Acceptable (5-6 marks):** Adequate MDP formulation with some components missing. Basic state-action dynamics defined; learning strategy needs improvement. **Poor (3-4 marks):** Limited MDP formulation; key components absent. Limited state-action dynamics; learning strategy lacking depth. **Very poor (0-2 marks):** Inadequate MDP formulation; components unclear. Inadequate state-action dynamics and learning strategy.	**10 Marks**
Discussion and interpretation **Outstanding (9-10 marks):** Profound insights derived; comprehensive analysis presented. **Good (7-8 marks):** Valuable insights drawn with thorough analysis. **Acceptable (5-6 marks):** Meaningful insights highlighted; analysis is sufficient. **Poor (3-4 marks):** Basic insights provided; analysis lacks depth. **Very poor (0-2 marks):** Insights minimal or missing; analysis insufficient.	**10 Marks**
Project report quality and organisation **Outstanding (9-10 marks):** Exceptional organisation with clear sections and logical flow. **Good (7-8 marks):** Well-structured report with coherent sections. **Acceptable (5-6 marks):** Adequate report structure with sections mostly organised. **Poor (3-4 marks):** Basic organisation with some inconsistencies. **Very poor (0-2 marks):** Poorly organised report: sections lack coherence.	**10 Marks**