Assessment 2: Cloud Computation

Due: *11:59 PM AEST Sunday week 4*

Weight: *4 0%*

Overview

During this assessment, you will produce a written report on neural network using AWS SageMaker.

Learning outcomes

1. Analyse real world tasks using machine learning methods, in particular describing, choosing, and applying appropriate supervised machine learning methods for descriptive data mining tasks
2. Develop and deploy machine learning models on AWS
3. Tune hyperparameters for machine learning models using AWS
4. Programmatically interact with AWS using Python Jupyter Notebooks
5. Synthesise and communicate the method and findings to diverse audiences

Format

You will need to submit the following:

- A PDF file clearly shows the assignment question, the associated answers, relevant Python outputs, analyses and discussions.

- Appendices include Jupyter Notebook or Python code or screen images of AWS console detailing the procedure of building, training and building of the NN model.

- The assignment **should not exceed 15-A4 pages.** Appendices do not form part of the page limit.

- The task cover sheet

You have up to three attempts to submit your assessment, and only the last submission will be graded.

A word on plagiarism

Plagiarism is the act of using another's words, works or ideas from any source as one's own.

Plagiarism has no place in a University. Student work containing plagiarised material will be subject to formal university processes.

Background

Understanding the possibility of readmission of diabetes patients may provide hospitals and doctors insight into effectiveness of ongoing treatments, and potential changes in the treatments. The changes in treatments may potentially save patients' life.

You are asked to investigate if a neural network accurately classifies if a patient with diabetes is likely to be readmitted to the hospital given their current conditions.

Data
The data, **"diabetic_data.csv"**, contains 10 years (1999-2008) of clinical care at 130 U.S. hospitals and integrated delivery networks. Data description is provided in Table 1, *"Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records"* BioMed Research International, vol. 2014, Article ID 781670, https://www.hindawi.com/journals/bmri/2014/781670/

The data can be downloaded from learnJCU and is from https://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008

Assessment Tasks

1. **Prepare** data appropriate for the proposed neural
   a) Discuss a subset of relevant predictors used in a NN model
   b) Apply and Discuss any appropriate cleaning or transformations
   c) Apply and discuss the training and testing dataset.

2. **Build, train and deploy a neural network**
   a) Propose a neural network for the classification/prediction. Discuss the structure of your proposed neural network and explain the total number of parameters required in the neural network.
   b) Build and train the proposed model in AWS SageMaker. Justify your choices of loss function, parameters and hyperparameters of the model
   c) Report and discuss the performance(s) and interpretations(s) of the chosen model
   d) Apply techniques such as dropout, early stop, batch normalisation to the benchmark model, and investigate their impacts on the performance of the neural network model.
   e) Provide suggestions / Discuss limitations of the proposed model in addressing the classification task.
   f) Provide evidence of endpoints and deploying all models in the AWS Sagemaker

3. **Discuss** the considerations of using the AWS SageMaker. At least include discussion regarding:
   a) Notebook instance type
   b) Cost and computation time

Marking Criteria and Rubric: MA5852 Assessment 2

Criteria	High Distinction	Pass	Fail
Prepare data for analysis **10% of total grade**	Demonstrate superior ability to clean and prepare data for ML and NN analysis. Highly developed awareness of the processes needed to prepare data for analysis	Demonstrate limited ability to clean and prepare data for ML and NN analysis. Limited awareness of the processes needed to prepare data.	Demonstrate poor or no ability to clean and prepare data for ML and NN analysis. No awareness of the processes needed to prepare data.
Build, train and deploy neural network **40% of total grade**	Demonstrate superior ability to justify and design a structure of NN, the training and testing dataset, select hyperparameters for NN to analyse data. Demonstrate superior ability to logically arrange, present and communicate the information of analysis and comparison.	Demonstrate limited ability to design a NN, the training and testing dataset, select hyperparameters for NN to analyse data. Demonstrate limited ability to logically arrange, present and communicate the information of analysis and comparison.	Demonstrate poor or no ability to design a NN, investigate and select hyperparameters for NN to analyse data. Poor to no ability to logically arrange, present and communicate the information of analysis and comparison.
Regularisation **30% of total grade**	Provide detailed discussion regarding impacts of regularisation approaches and batch normalisation on the performance of the model. Provide sound suggestions to further improve the	Provide limited discussion regarding impacts of regularisation approaches and batch normalisation on the performance of the model. Provide some suggestions to further improve the	Provide poor discussion regarding impacts of regularisation approaches and batch normalisation on the performance of the model. Provide some suggestions to further improve the

	performance of the model.	performance of the model.	performance of the model.
Implementing NN on AWS 20% of total grade	Demonstrate superior ability to build, train and deploy NN in AWS SageMaker.	Limited ability to build, train and deploy NN in AWS SageMaker.	Poor to no ability to build, train and deploy NN in AWS SageMaker.