Sample - Assessment 2

Instructions

You will have a duration of 24 hours to work on this assessment.

1. Note **this is a practice assessment**, the actual one will be given end of **Week 5 (Friday 5 pm Sydney time).**
2. You cannot do this in a group and must ensure that you do not copy work from others and the internet - cite relevant sources when you refer to online material.
3. You need to **choose one** option.

Option I: Model Development and Report

Use this Dataset for this question: https://archive.ics.uci.edu/ml/datasets/ionosphere

ionosphere.data and ionosphere.names are the key files

Task-1 (4 marks) -Data processing

Data processing for machine learning: Given the attached data set, use either R or Python with the needed libraries to process the data.

- Use either code or your text editor and run a simple find and replace to transform the text in the data to numbers. Normalise the input features between 0 and 1 and use one-hot-encoding for the target classes. Save normalised data as data.csv (using appropriate code).(1 Mark)
- Show distribution of features using data appropriate visualisation tools (histogram/boxplots) (1 Marks)
- Provide further visualization of the dataset using a heatmap of the feature correlation matrix. (1 Marks)
- Write a paragraph about major observations in your report (1 Mark).

Task 2 - (8 marks) - Model

You can use either R or Python with the needed libraries for this task.

Apply neural network (with the appropriate number of hidden layer and neurons) using 50/50 percent train and test split. At least 10 experimental runs (either with train/test split of data or different initial positions in weight space for model) are required for each task.

- Compare Adam with SGD, report performance of the training and test data set (percentage correctly classified) (2 Marks)
- Apply L2 regularisation (weight decay) and compare results of Adam and SGD with and without regularisation. (2 Marks)
- Show the confusion matrix and AUC/ROC curve **from a single experimental run** (2 Marks)
- Describe and interpret your results. (1 Mark)
- Discuss limitations and future directions of how this work can be extended (1 Mark)

Task 3 - 8 marks - Technical Report

Create a technical report and upload pdf.

- You should ensure that your results are provided in a Table - especially when you are comparing methods. You need to ensure that references are formatted to at least two significant figures.
- Tables and Figures need to be numbered and cited in-text when you are discussing or introducing them.
- Ensure that Figures are visible and placed in the appropriate order.
- Ensure that you have a brief introduction and conclusion section in your report.
- Do not include code and screenshots of results from software in your report.

Resources

Assesment 1 sample solutions: https://github.com/rohitash-chandra/dataminingMATH5836/tree/master/assesments/assesment1/solutions

Report Rubrics:

1. The paper has an excellent presentation. The introduction clearly defines the aim and goals of the paper. The results and discussion section has been presented very well. Some references are cited. (8 Marks)
2. The paper has a good presentation. The introduction clearly defines the aim and goals of the paper. The results and discussion section has been presented well but some issues are present. (6 Marks)
3. The paper has some presentation issues. The introduction does not clearly define the aim and goals of the paper. The results and discussion section has not been presented very well. (4 Marks)
4. The paper has a poor presentation. The introduction has missing aim and goals. The results and discussion section is questionable or not complete. (2 Marks)
5. No submission/results not correct (0 Marks)

Option II - Research Essays

Option II: Essay in data science

- Discuss what are the key similarities and differences between neural networks and logistic regression (6 Marks).
- Discuss the similarities and differences between SGD and Adam optimiser (6 Marks).
- Discuss the major limitations of neural networks and how Bayesian inference can be useful (8 Marks).

Limit each to around 400 - 500 words and cite references (APA style) to support your viewpoint. Upload pdf of the report.

You should cite the references in-text: https://guides.lib.monash.edu/citing-referencing/APA-In-texthttps://owl.purdue.edu/owl/research_and_citation/apa_style/apa_formatting_and_style_guide/in_text_citations_the_basics.html

Resources:

- https://apastyle.apa.org/style-grammar-guidelines/references/examples