MS4S10 - 2020/21

Assessment 2

Deadline: Tuesday 16th March 2020, 9:00 PM

This assessment is worth 50% of your overall mark for this module.

You must use Python to perform your analyses and produce a Jupyter notebook summarising your output and results. Your findings should be interpreted, and valid conclusions drawn within the notebook.

Only ONE file, which includes the assessment cover sheet and notebook, is to be submitted to Turnitin before the deadline, with contents converted to a PDF format.

Note: Send a copy of the ipynb file filippo.cavallari@southwales.ac.uk before the deadline. The subject of the email MUST be: "MS4S10 CW2 Jupyter Notebook".

Task 1 – Gradient Descent (15%)

Apply the gradient descent algorithm to find the minimum of the following function of four variables

$$f(x, y, z, w) = \frac{1}{4}(2 - x)^2 + (3y - 5)^4 + e^{2z^4 + w^2}$$

Comment on the results.

Task 2 – Logistic Regression as a Neural Network with another activation function (45%)

Write a report on how to consider Logistic Regression as a Neural Network with one single neuron using

$$\tanh z = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$

as the activation function.

Observe that the range of the hyperbolic tangent is the interval $(-1; 1)$. You will need to rescale the activation function and/or the cross-entropy loss function as required.

Focus in particular on:

1. The calculation of the derivative of the activation function and how to express the derivative using the hyperbolic tangent itself.

2. Discuss the behaviour of the cross-entropy loss function. Use matplotlib to draw appropriate graphs and comment on the results.
3. Evaluate

$$\frac{\partial \mathcal{L}^{(i)}}{\partial w_k} \quad and \quad \frac{\partial \mathcal{L}^{(i)}}{\partial b}.$$

4. Write a Python script to simulate this neuron and use it to classify a range of different examples chosen by you
5. Does this neuron have a better performance than the neuron created in class using the logistic function as the activation function? Explain your reasoning.

Provide as many mathematical details as possible to show your workings.

Task 3 – MNIST (30%)

Consider the example of the MNIST digits datasets we discussed in class. Try at least 4 different Neural Networks architectures (by changing the number of neurons and the number of layers) and compare their performances using:

1. At least two activation functions (sigmoid, relu, …);
2. At least four different step sizes;
3. At least three batch sizes;
4. At least three optimisation algorithms (remember to use at least three different values for the momentum when using gradient descent);
5. At least two parameter initialisations.

Draw appropriate graphs and comment on them to support your conclusions.

Jupyter Notebook (10%)

Your code should be clear, efficient and well commented. The notebook must be well structured, well written and comments should be grammatically correct. Titles, heading and figures should be correct and labelled in a meaningful way and referenced accordingly.

Marking Guidelines

	80-100	70-79	60-69	50-59	40-49	30-39	0-29
	Exceptional First	**First**	**Upper 2nd**	**Lower 2nd**	**Third**	**Narrow Fail**	**Fail**
Analysis outline	Professional outline of analysis presented.	Detailed purpose of analysis provided.	Adequate outline of analysis provided.	Outline of analysis provided but with some flaws.	Simple outline of analysis provided, but lacking key detail.	Inadequate outline of analysis provided.	No outline of analysis provided.
Methods used and assumptions made	Sophisticated investigation of assumptions and description of methods used.	Comprehensive investigation of assumptions and description of methods used.	Adequate and correct investigation of assumptions and description of methods used.	Investigation of assumptions and description of methods used is provided but with some flaws.	Limited investigation of assumptions and description of methods used.	Inadequate investigation of assumptions and description of methods used.	No investigation into assumption and no description of methods used.
Neural Network Analysis	Sophisticated techniques utilised.	Comprehensive techniques utilised	Adequate techniques utilised.	Techniques attempted but with some flaws.	Limited techniques utilised.	Inadequate techniques utilised.	No correct techniques utilised.
Key results and correctness of content	Unanticipated results and implementations presented. Appropriate, substantial, correct and sophisticated nature.	Comprehensive results and implementations, presented and employed well. Appropriate, substantial and correct.	Expected results and implementations presented. All appropriate, largely correct, with few flaws.	Not all expected results and implementations presented. All appropriate, largely correct, with few flaws	Few or simple results and implementations presented. Much appropriate material, but flawed.	Seriously flawed results or no implementation. Appropriate but seriously flawed material.	No results or implementation. Incorrect or inappropriate content.
Conclusions	Deep and critical understanding provided.	Thorough understanding shown.	Good understanding shown.	Key concepts generally understood.	Some evidence of understanding.	Little of superficial understanding shown.	No evidence of understanding.
Jupyter notebook	Like a publishable report, virtually error-free.	Like a publishable report with isolated minor errors.	Can be followed easily with very few errors.	Can be followed easily with some weaknesses.	Can be followed with difficulty.	Poor structure or containing significant errors.	Unstructured and with many errors.