Problem Set 5

Advanced data analysis and machine learning

> **Important**
>
> You must upload your problem set to Moodle before 23:55 on 2023-06-02.
>
> Your submission must include:
>
> - a PDF that responds to all questions; and
> - a single Python script (`.py`) that you used to answer all questions.
>
> Answers without accompanying Python code will recieve zero marks.

Question 1 [40 points]

In this file you will find a set of 100 observations. Each observation has some features x_1 and x_2, and some classification of 0 or 1. Plot the data x_1 and x_2, coloured by the classification c. Your task is to build a neural network that will predict some classification given x_1 and x_2. Your network should have one hidden layer, with seven neurons in the hidden layer. Chose an activation function. Draw the network, labelling inputs and weights. Derive the updated estimates for the weights by finding the derivatives of the loss function with respect to the weights.

Question 2 [30 points]

Implement the neural network from Question 1 using `numpy`. Train the neural network using an appropriate portion of the data, and plot the training and test loss as a function of epoch. Plot a smooth map of classifications in x_1 and x_2, and overlay the data.

Note that you need to make appropriate decisions regarding learning rates, initialisation, data segmentation, optimizer, number of epochs, etc. You should justify these decisions.

Question 3 [20 points]

Implement the network from Question 1 using an off-the-shelf machine learning library of your choice (in any language). Justify your decisions as per Question 2, and make the same plots as per Question 2. Comment on any differences.

Question 4 [10 points]

Explain the following terms:

- activation function
- back propagation
- learning rate
- vanishing gradients
- weights and biases
- loss (training, validation, and test)
- convolution kernels
- pooling

- data augmentation
- softmax layers