PHYS/4036 Workshop 2

Adam Moss and Maggie Lieu

January 2020

Question 1:

For the linear model, given by $\hat{\mathbf{Y}} = \mathbf{X}\mathbf{w}$, derive the normal set of equation for the optimal weights, using an MSE loss with an additional L_2 penalty term of the form $\frac{\lambda}{2}||\mathbf{w}||_2^2$.

Question 2:

In this question you are given a dataset (MNIST) with a limited number of training examples (only 1000 compared to the usual 60,000).

Your goal is to implement regularisation methods to achieve the lowest possible **test loss** using this dataset. You should consider methods given in the lectures including:

1. Data augmentation
2. Early stopping
3. L1/L2 penalty norms
4. Dropout

You are free to change the network architecture and model complexity, but the main purpose of the workshop is to investigate regularisation (next week you will look at CNN architectures in detail). You are also free to change the choice of optimiser, and other hyper-parameters such as the batch size.

The starting notebook from which to implement regularisation methods is:

```
https://github.com/adammoss/MLiS2/blob/master/workshops/
workshop2/regularisation.ipynb
```

In your submitted notebook, please do not change the test/train split.