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 $\lambda||\mathbf{w}||_2^2$.

Question 2:

The pre-activation value of a neuron i in layer ℓ is

$$z_i^{(\ell)} = \sum_{j=1}^{n_{\text{inputs}}} a_j^{(\ell-1)} W_{ji}^{(\ell)} \,. \tag{1}$$

Assuming the activations and weights are independent and identically distributed with zero mean, and the weights are uniformly distributed with $W_{ij} \sim \mathcal{U}(-\alpha, \alpha)$, what is the value of α required for the pre-activation $z_i^{(\ell)}$ to have the same variance as each of the activations $a_j^{(\ell-1)}$?

[Hint – you will need to calculate the variance of $\mathcal{U}(-\alpha, \alpha)$. You do not need to consider balancing the variance of the backwards pass.]

Question 3:

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.