PHYS/4036 Workshop 1

Adam Moss

January 2020

Question 1:

For the logistic regression model with a binary cross-entropy (BCE) loss, show that the derivative of the loss function with respect to the weights is given by

$$\nabla_{\mathbf{w}} J(\mathbf{w}) = \frac{1}{N} \sum_{i}^{N} \left(\hat{y}^{(i)} - y^{(i)} \right) \mathbf{x}^{(i)} = \frac{1}{N} \mathbf{X}^T \left(\hat{\mathbf{Y}} - \mathbf{Y} \right) . \tag{1}$$

Question 2:

Modify the back-propagation example given in lectures:

```
https://github.com/adammoss/MLiS2/blob/master/examples/intro/backprop.ipynb
```

to use a sigmoid activation function at the output layer and a BCE loss. Train the model using back-propagation to learn the AND and XOR functions.

Question 3:

In lectures we wrote down the back-propagation equations for a multi-layer perceptron consisting of an input layer with two units, a hidden layer with two hidden units, and a single output unit.

Derive equivalent equations to (1.32-1.34 and 1.36) for the same network, but now with two hidden layers, each with 2 hidden units. You should write down the error of each neuron $\Delta_j^{(\ell)}$ and the derivative of the loss with respect to the weights $\partial J / \partial W_{kj}^{(l)}$.

Question 4:

Implement the back-propagation equations for the network in Q3 to learn the AND and XOR functions, and investigate the convergence of your algorithm with respect to the choice of initial weights and learning rate.