1 Preamble

The neural network (function approximator) is just a chain of geometric transformations (functions) each parametrized by a weight matrix $W \in \mathcal{R}^{n_x \times n_h}$ and a bias vector $b \in \mathcal{R}^{n_h}$ on the input matrix $X \in \mathcal{R}^{m \times n_x}$. Note that m is the number of examples in the dataset, n_x is the input size (i.e., output of the previous layer), and n_h is number of hidden units in the current layer. Also θ refers to the weights W and biases b in the network. Learning occurs by updating the parameters iteratively via mini-batch (stochastic) gradient descent.

2 Equations

Here I define the operations that occur for a single hidden layer fully connected neural network (multilayer perceptron or MLP) with the activation function of the last layer as the sigmoid function $\sigma(\cdot)$ during training. **POSSIBLY NOT SIGMA?**

2.1 Single Hidden Layer Neural Network

$$\begin{aligned} NeuralNet_\theta(X) &= \sigma(ReLU(XW^{[1]} + b^{[1]})W^{[2]} + b^{[2]}) \\ &= \sigma(g(ReLU(w(X)))) \\ A^L &= NeuralNet_\theta(X) \qquad \text{Activation matrix } A \text{ for last layer } L \end{aligned} \tag{1}$$

where σ, $ReLU$, g, and w are defined as follows:

$$\begin{aligned} \sigma(t) &= \frac{1}{1+e^{-t}} \\ ReLU(t) &= max(0, t) \\ g_{\theta^{[2]}}(A) &= AW^{[2]} + b^{[2]} \\ w_{\theta^{[1]}}(A) &= AW^{[1]} + b^{[1]} \\ u_{\theta^{[l]}}(A) &= AW^{[l]} + b^{[l]} \qquad \text{General form of } g \text{ and } w \text{ for } l^{th} \text{ layer} \end{aligned} \tag{2}$$

2.2 Neural Network Prediction (Forward Pass)

$$\hat{y} \leftarrow NeuralNet_\theta(X) \tag{3}$$

2.3 Mean Squared Error Loss Function

$$\begin{aligned}\mathcal{L}_\theta(X) &= \frac{1}{N}\sum_{i=1}^{N}(\hat{y}^{(i)} - y^{(i)})^2 \\ &= \frac{1}{N}\sum_{i=1}^{N}(a^{(i)} + y^{(i)})^2 \quad a^{(i)} \text{ is the activation vector of the last layer } L \text{ for the } i^{th} \text{ input}\end{aligned} \tag{4}$$

2.4 Gradient Update

$$\theta_i \leftarrow \theta_i - \eta(\nabla_\theta \mathcal{L}_\theta(\hat{y}, y)) \tag{5}$$