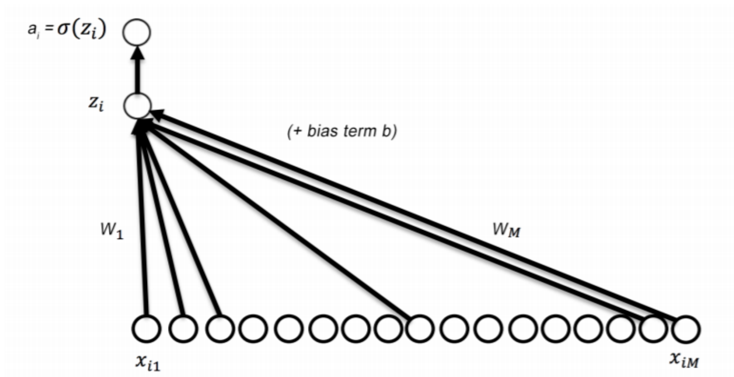
CEE690/ECE590: Homework 1 Solutions

1 Problem 1

1.1 Part A



Model set-up:

Pre-activation: $z = \sum_{j=1}^{M} w_j x_j + b$
Activation: $a = \sigma(z) = \frac{1}{1+e^{-z}}$
Loss: $\mathcal{L} = -\Big(y \log(a) + (1-y)\log(1-a)\Big)$

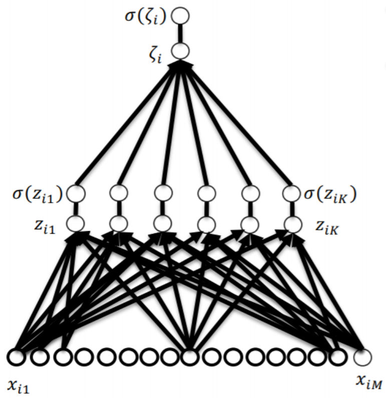
Gradients:

Weights: $\frac{\partial \mathcal{L}}{\partial w_j} = \frac{\partial \mathcal{L}}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial w_j}$
Biases: $\frac{\partial \mathcal{L}}{\partial b} = \frac{\partial \mathcal{L}}{\partial a}\frac{\partial a}{\partial z}\frac{\partial z}{\partial b}$

with $\frac{\partial \mathcal{L}}{\partial a}, \frac{\partial a}{\partial z}, \frac{\partial z}{\partial w_j}, \frac{\partial z}{\partial b}$ defined as the following:

$\frac{\partial \mathcal{L}}{\partial a} = -\frac{y}{a} + \frac{1-y}{1-a}$
$\frac{\partial a}{\partial z} = \frac{e^{-z}}{(1+e^{-z})^2} = a(1-a)$
$\frac{\partial z}{\partial w_j} = x_j$
$\frac{\partial z}{\partial b} = 1$

1.2 Part B



Model set-up:

Layer 1:

Pre-activation: $z_k^{(1)} = \sum_{j=1}^{M} w_{kj}^{(1)} x_j + b^{(1)}$

Activation: $a_k^{(1)} = \sigma(z_k^{(1)}) = \frac{1}{1+e^{-z_k^{(1)}}}$

Layer 2:

Pre-activation: $z^{(2)} = \sum_{k=1}^{M} w_k^{(2)} a_k + b^{(2)}$

Activation: $a^{(2)} = \sigma(z^{(2)}) = \frac{1}{1+e^{-z^{(2)}}}$

Loss: $\mathcal{L} = -\Big(y \log(a^{(2)}) + (1-y)\log(1-a^{(2)})\Big)$

Gradients:

Weights:

$\frac{\partial \mathcal{L}}{\partial w_k^{(2)}} = \frac{\partial \mathcal{L}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial z^{(2)}} \frac{\partial z^{(2)}}{\partial w_k^{(2)}}$

$\frac{\partial \mathcal{L}}{\partial w_{kj}^{(1)}} = \frac{\partial \mathcal{L}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial z^{(2)}} \frac{\partial z^{(2)}}{\partial a_\kappa^{(1)}} \frac{\partial a_\kappa^{(1)}}{\partial z_k^{(1)}} \frac{\partial z_k^{(1)}}{\partial w_{kj}^{(1)}}$

Biases:

$\frac{\partial \mathcal{L}}{\partial b^{(2)}} = \frac{\partial \mathcal{L}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial z^{(2)}} \frac{\partial z^{(2)}}{\partial b^{(2)}}$

$\frac{\partial \mathcal{L}}{\partial b^{(1)}} = \frac{\partial \mathcal{L}}{\partial a^{(2)}} \frac{\partial a^{(2)}}{\partial z^{(2)}} \frac{\partial z^{(2)}}{\partial a_\kappa^{(1)}} \frac{\partial a_\kappa^{(1)}}{\partial z_k^{(1)}} \frac{\partial z_k^{(1)}}{\partial b^{(1)}}$

Terms in chain rule defined similarly as in 1a)

1.3 Part C

To change 1a) to multi-class, we replace the vector dot product with a matrix multiply and replace the binary cross entropy with a softmax cross entropy:

Model set-up:

Pre-activation: $z_k = \sum_{j=1}^{M} w_{kj} x_j + b$
Activation: $a_k = \frac{e^{z_k}}{\sum_\kappa e^{z_\kappa}}$
Loss: $\mathcal{L} = -\sum_k y_k \log(a_k)$

Similar changes are made to the top layer 1b).

The gradients for 1a) and 1b) remain the same, with the gradients of the new layers being:

Gradients:

$\frac{\partial \mathcal{L}}{\partial a_k} = -\sum_k \frac{y_k}{a_k}$

$\frac{\partial a_k}{\partial z_j} = \begin{cases} a_k(1 - a_k), & \text{if } k = j \\ -a_j a_k, & \text{otherwise} \end{cases}$