Assume a two layer feedforward MLP with a single output where all the elements use sigmoid activation (including the output unit). Assume a training dataset {x(i), y(i)}mi = 1 where x(i) ∈ n and y(i) ∈ {0, 1}. Derive the backpropagation update equations for the weights of the output by minimizing the following error function: $\frac{1}{2}\sum_{i=1}^{m}(y^{(i)}-\hat{y}^{(i)})^2$ Compare the result you obtain to that obtained in class using maximum likelihood estimation.

**My update equations for the weights of the output by minimizing the following error function:**

$$\frac{1}{2}\sum_{i=1}^{m}(y^{(i)}-\hat{y}^{(i)})^2$$

The neural network model has unknown parameters, often called weights, and we seek values for them. We denote the complete set of weights by Θ which consists of

$$\{\alpha_{0m},\alpha_m;m=1,2,...M\} \quad \text{M(P+1) weights}$$

$$\{\beta_{0k},\beta_k;k=1,2,...,k\} \quad \text{K(M+1) weights}$$

Depends

$$R(\theta)=\sum_{i=1}^{N}R_i=\sum_{i=1}^{N}\sum_{k=1}^{K}(y_{ik}-f_k(x_i))^2$$

$$\frac{\partial R_i}{\partial \beta_{km}}=-2(y_{ik}-f_k(x_i))g_k^{'}(\beta_k^T z_i)z_{mi}$$

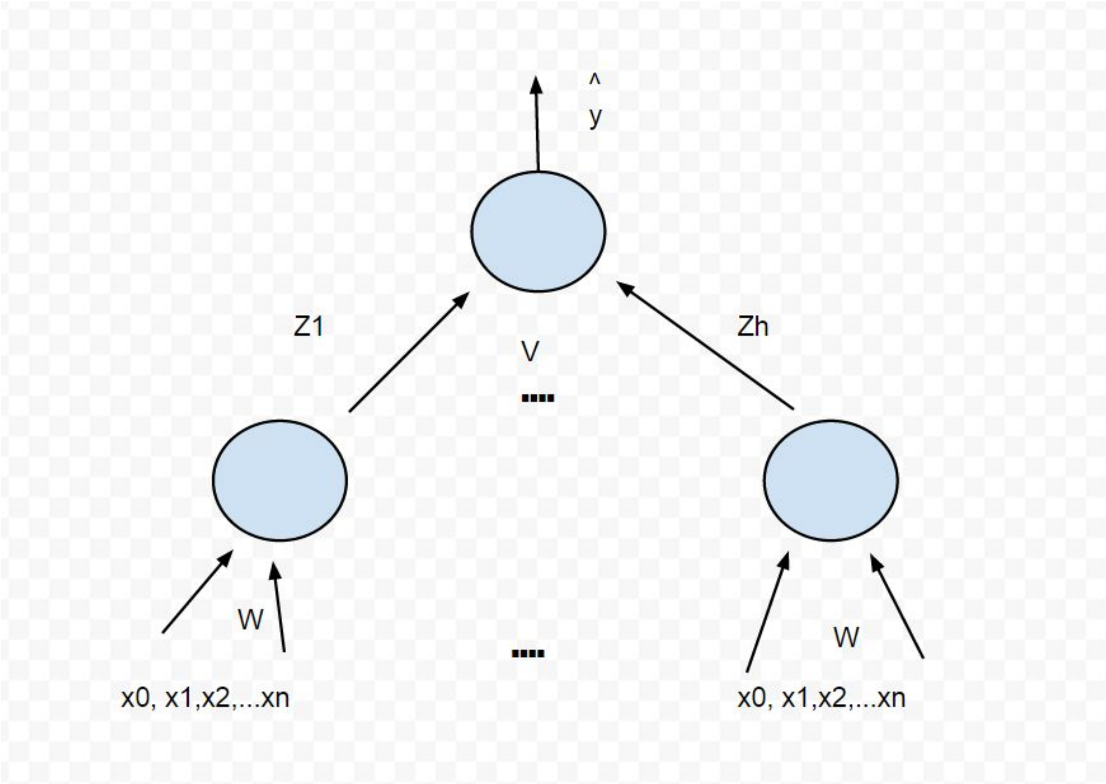
$$\frac{\partial R_i}{\partial \alpha_{ml}}=-\sum_{k=1}^{k}2(y_{ik}-f_k(x_i))g_k^{'}(\beta_k^T z_i)\beta_{km}\sigma^{'}(\alpha_m^T x_i)x_{il}$$

Gradient Decent:

$$\beta_{km}^{(r+1)}=\beta_{km}^{(r)}-\sum_{i=1}^{N}\frac{\partial R_i}{\partial \beta_{km}^{(r)}}$$

$$\alpha_{ml}^{(r+1)}=\alpha_{ml}^{(r)}-\sum_{i=1}^{N}\frac{\partial R_i}{\partial \alpha_{ml}^{(r)}}$$

**The Maximum likelihood estimation**




Objective:

$$E(\{w_i\},V) = \frac{1}{2}\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})^2$$

Update parameters:

$$v_{i+1} \leftarrow v_i - \eta\frac{\partial E}{\partial V}$$

$$w_{j+1} \leftarrow w_j - \eta\frac{\partial E}{\partial w_j}$$

Using chain rule:

$$\frac{\partial E}{\partial V} = \frac{\partial E}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial V}$$

$$\frac{\partial E}{\partial w_j} = \frac{\partial E}{\partial \hat{y}}\frac{\partial \hat{y}}{\partial z_j}\frac{\partial z_j}{\partial w_j}$$

We get

$$v \leftarrow v - \eta\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})z^{(i)}$$

$$w_j \leftarrow w_j - \eta\sum_{i=1}^{m}(\hat{y}^{(i)} - y^{(i)})v_j z_j^{(i)}(1 - z_j^{(i)})x^{(i)}$$