HOMEWORK # 4 | ELEG 867 | Max Sokolich

QUESTION # 1

11.9 $R(\theta) = \sum_{1}^{K} \sum_{1}^{N} (y_{ik} - f_K(x_i))^2$

recall

$J(\theta) = \frac{1}{2} \sum (a_i^L - y_i)^2$

$a^L = h_\theta(x) = g(z^L) = f_K(x_i)$

11.10: $R(\theta) = - \sum_{i=1}^{N} \sum_{k=1}^{K} y_{ik} \log f_K(x_i) \longrightarrow$ form forward back equations

11.3

let $z_{mi} = \sigma(\alpha_{0m} + \alpha_m^T x_i)$ with $z_i = (z_{1i}, z_{2i} \ldots z_{Mi})$

$$\frac{\partial R_i}{\partial \beta_{km}} = \underbrace{- \frac{y_{ik}}{f_K(x_i)} g'_K(\beta_K^T z_i)}_{\delta_{ki}} z_{mi} \qquad \text{(1)}$$

$$\frac{\partial R_i}{\partial \alpha_{m\ell}} = \underbrace{- \sum_{k=1}^{K} \frac{y_{ik}}{f_K(x_i)} g'_K(\beta_K^T z_i) \beta_{km} \sigma'(\alpha_{0m} + \alpha_m^T x_i)}_{s_{mi}} x_{i\ell} \qquad \text{(2)}$$

with (1) and (2)

we can find

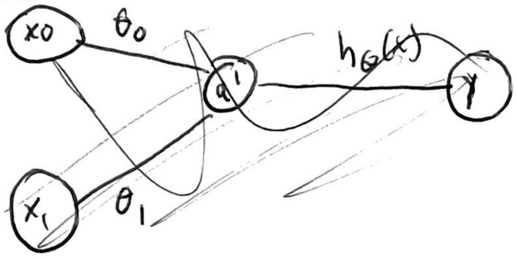
$$\beta_{km}^{(r+1)} = \beta_{km}^{(r)} - \gamma_r \sum_{i=1}^{N} \frac{\partial R_i}{\partial \beta_{km}^r}$$

$$\alpha_{m\ell}^{r+1} = \alpha_{m\ell}^r - \gamma_r \sum_{i=1}^{N} \frac{\partial R_i}{\partial \alpha_{m\ell}^r}$$

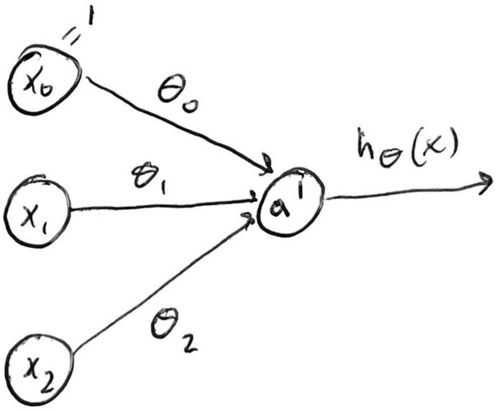
$$\boxed{s_{mi} = \sigma'(\alpha_{0m} + \alpha_m^T x_i) \sum_{1}^{K} \beta_{km} \delta_{ki}}$$

equation

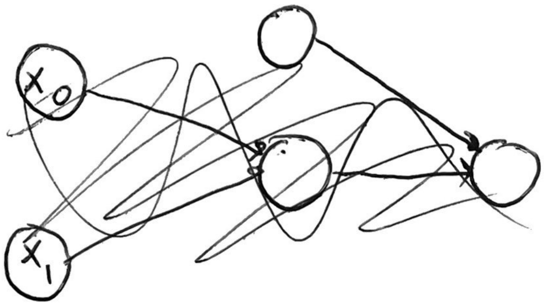
QUESTION # 2



$$h_\theta(\vec{x}) = g(\vec{\theta}, \vec{x}) = \begin{cases} 1 & x > 0 \\ 0 & \text{else} \end{cases}$$



let $\theta_1 = \theta_2 = 1.5$, $\theta_0 = -2.0$



a) $\vec{x}^{(1)} = (0, 0)$

$$a^1 = h_\theta(x) = y = g(\theta_0 + x_1\theta_1 + x_2\theta_2) = g(-2 + 0 \cdot 1.5 + 0 \cdot 1.5)$$

$$= g(-2) = \boxed{0}$$

b) $x^2 = (1, 0)$

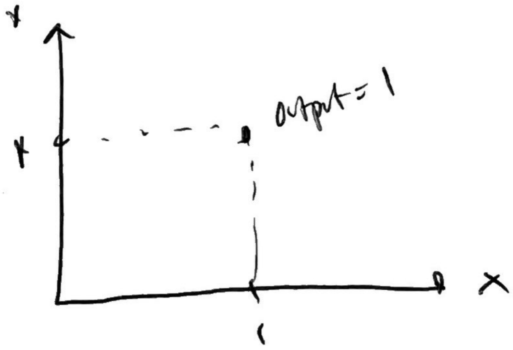
$$y = g(-2 + 1(1.5) + 0(1.5)) = g(-0.5) = \boxed{0}$$

c) $x^3 = (0, 1)$

$$y = g(-2 + 0(1.5) + 1(1.5)) = g(-0.5) = \boxed{0}$$

d) $x^4 = (1, 1)$

$$y = g(-2 + 1.5 + 1.5) = g(1) = \boxed{1}$$



truth table

x_1	x_2	y
0	0	0
0	1	0
1	0	0
1	1	1

AND GATE

QUESTION 3

a) perceptron Not Gate

$$y_1 = g(\theta_{10} + x\theta_{11} + x_2\theta_{12})$$
$$y_2 = g(\theta_{20} + x_1\theta_{21} + x_2\theta_{22})$$

$$\theta_{12} = \theta_{21} = 0$$

$$y_1 = g(\theta_{10} + x_1\theta_{11})$$
$$y_2 = g(\theta_{20} + x_2\theta_{22})$$

$$\vec{\theta} = \begin{bmatrix} \theta_{10} & \theta_{11} \\ \theta_{20} & \theta_{22} \end{bmatrix}$$

let

$$g(v) = \begin{cases} 1 & v \geq 0 \\ 0 & \text{else} \end{cases}$$

truth table

x_1	x_2	y_1	y_2
0	0	1	1
0	1	1	0
1	0	0	1
1	1	0	0

4 equations

$$1 = g(\theta_{10} + 0\,\theta_{11})$$
$$1 = g(\theta_{20} + 0\,\theta_{22})$$
$$0 = g(\theta_{10} + \theta_{11})$$
$$0 = g(\theta_{20} + \theta_{22})$$

$$1 = g(\theta_{10})$$
$$1 = g(\theta_{20})$$
$$0 = g(\theta_{10} + \theta_{11})$$
$$0 = g(\theta_{20} + \theta_{22})$$

$$\boxed{\begin{aligned} \theta_{10} &\geq 1 \\ \theta_{20} &\geq 1 \\ \theta_{11} &< -1 \\ \theta_{22} &< -1 \end{aligned}}$$

b) and gate is just QUESTION 2

$$\boxed{\begin{aligned} \theta_1 = \theta_2 &= 1.5 \\ \theta_0 &= -2.0 \end{aligned}}$$

QUESTION #4 Activation functions

1) Binary step function

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

↳ considered the simplest activation function

↳ not useful with multiple classes

↳ gradient is zero causes issues in back prop

2) Tanh

$$\tanh(x) = 2\,\text{sigmoid}(2x) - 1$$

↳ sometimes avoided due to vanishing gradient problem

↳ output $(-1, 1)$

→ symmetric about the origin

↳ inputs of next layers will not always be of the same sign

3) ReLU

$$f(x) = \max(0, x)$$

↳ does not activate all neurons at the same time

↳ far more computationally efficient since only a certain amount of neurons are activated at a time

MOST USED

↳ gradient is zero → issues in back prop

4) SWISH

$$f(x) = x \text{ sigmoid}(x)$$

↳ differentiable at all points

↳ not monotonic

5) softmax

$$\sigma(z)_j = \frac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}, \quad j = 1, K$$

↳ described as a combination of multiple sigmoids

↳ good for multiclass problems