Theory Questions

Sri Hari Malla - CS19BTECH11039

October 29, 2021

1 Neural Networks

1.1

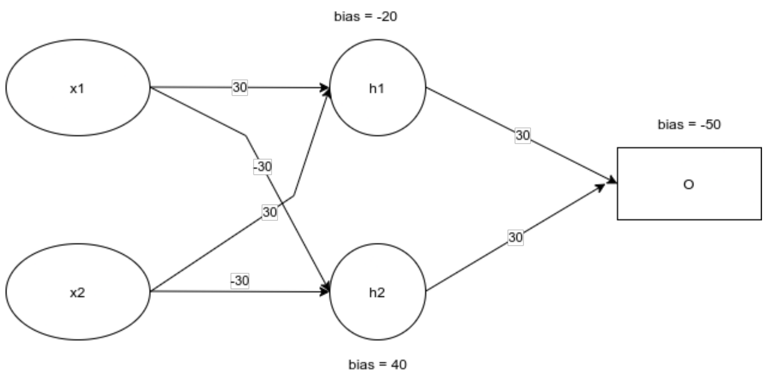

Figure 1: Network Representation

- **case 1** : x1=1 x2=1
  h1 = 30+30-20 = 40 >0 so h1=1
  h2 = -30-30+40 = -20 <0 so h2=0
  y1= 30-50 = -20 <0 so answer is 0.

- **case 2**: x1=1 x2=0
  h1 = 30-20 = 10 >0 so h1=1
  h2 = -30+40 = 10 <0 so h2=1
  y1= 30+30-50 = 10 >0 so answer is 1.

- **case 3**: x1=0 x2=1
  h1 = 30-20 = 10 >0 so h1=1
  h2 = -30+40 = 10 >0 so h2=1
  y1 = >0 so answer is 1.

- **case 4**: x1=0 x2=0
  h1 = -20 < 0 so h1=0
  h2 = 40 >0 so h2=1
  y1= -50 <0 so answer is 0.

Hence, a two layer perceptron with one hidden layer can solve the XOR problem

1.2

Given f, q neurons with functions:

$$q = x - y$$
$$f = q * z$$

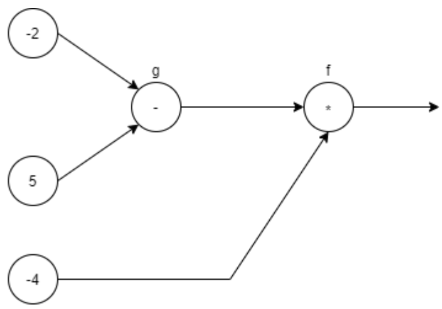

Figure 2: Graphical representation

Calculating the gradient of f with respect to x,y,z:

$$\frac{\partial f}{\partial x} = \frac{\partial (x-y)}{\partial x} * z = z = -4$$
$$\frac{\partial f}{\partial y} = \frac{\partial (x-y)}{\partial y} * z = -z = 4$$
$$\frac{\partial f}{\partial z} = \frac{\partial z}{\partial z} * (x-y) = x - y = -7$$

Sri Hari Malla
Indian Institute of Technology Hyderabad
Fundamentals of Machine Learning
CS19BTECH11039

2 Neural Networks

(2)

Given extension of cross-entropy error function for a multiclass problem is

$$E(w) = -\sum_{n=1}^{N}\sum_{k=1}^{K} t_{kn} \ln y_k(x_n, w)$$

Required to prove:

$$\frac{\partial E}{\partial a_k} = y_k - t_k$$

$$\frac{\partial E}{\partial a_k} = \frac{\partial}{\partial a_k}\left(-\sum_j t_j \ln(y_j)\right)$$

$$= -\sum_j t_j \frac{\partial}{\partial a_k}(\ln(y_j))$$

$$= -\sum_j t_j \cdot \frac{1}{y_j}\frac{\partial y_j}{\partial a_k}$$

$$\frac{\partial E}{\partial a_k} = -\frac{t_k}{y_k}\frac{\partial y_k}{\partial a_k} - \sum_{j!=k}^{K} t_j \frac{1}{y_j}\frac{\partial y_j}{\partial a_k} \longrightarrow (1)$$

Given $\quad y_k = \dfrac{\exp(a_k(x,w))}{\sum \exp(a_k(x,w))}$

$$\frac{\partial y_k}{\partial a_k} = \frac{\partial}{\partial a_k}\left(\frac{e^{a_k}}{e^{a_k} + \sum_{j!=k} e^{a_j}}\right)$$

$$= \frac{\left(e^{a_k} + \sum_{j!=k} e^{a_j}\right)e^{a_k} - e^{a_k}e^{a_k}}{\left(\sum_j e^{a_j}\right)^2}$$

$$= \frac{\left(\sum_{j} e^{a_j}\right)e^{a_k} - (e^{a_k})^2}{\left(\sum_j e^{a_j}\right)^2} = \frac{e^{a_k}}{\left(\sum_j e^{a_j}\right)} - \frac{(e^{a_k})^2}{\left(\sum_j e^{a_j}\right)^2}$$

$$\frac{\partial y_k}{\partial a_k} = y_k - y_k^2 \longrightarrow (2)$$

Figure 3: page1

Sri Hari Malla
Indian Institute of Technology Hyderabad
Fundamentals of Machine Learning
CS19BTECH11039

$$\frac{\partial y_j}{\partial a_k} = \frac{\partial}{\partial a_k}\left(\frac{e^{a_j}}{e^{a_k} + \sum_{j \neq k} e^{a_j}}\right)$$

$$= \frac{0 - e^{a_j} \cdot e^{a_k}}{\left(\sum_j e^{a_j}\right)^2}$$

$$\frac{\partial y_j}{\partial a_k} = -y_j \cdot y_k \longrightarrow (3)$$

Substituting (2) and (3) in (1)

$$\frac{\partial E}{\partial a_k} = -\frac{t_k}{y_k}\left(y_k - y_k^2\right) - \sum_{j \neq k}^{K} t_j \frac{1}{y_j}\left(-y_j y_k\right)$$

$$= -t_k(1 - y_k) + \sum_{j \neq k}^{K} t_j y_k$$

$$= -t_k + t_k y_k + \sum_{j \neq k}^{K} t_j y_k$$

$$\frac{\partial E}{\partial a_k} = -t_k + y_k\left(\sum_j t_j\right)$$

$$\sum_j t_j = 1 \Rightarrow \frac{\partial E}{\partial a_k} = y_k - t_k$$

Hence shown

Figure 4: page2

Sri Hari Malla
Indian Institute of Technology Hyderabad
Fundamentals of Machine Learning
CS19BTECH11039

3 Jensen's Inequality

For a real convex function ϕ numbers $x_1, x_2 \ldots \ldots x_n$ in it's domain and positive weights a_i, jensen's equality is written as:

$$\phi\left(\frac{\sum a_i x_i}{\sum a_i}\right) \leq \left(\frac{a_i \phi\left(x_i\right)}{\sum a_i}\right)$$

- Lets assume all weights as 1 so that summation of the weights is M.
- Given a convex function f(x)=$(x)^2$ and a convex function y.
- let's assume $x_i = [(y_m(x) - f(x))^2]$ and the real convex function ϕ is E_x.

$$E_{AV} = \frac{1}{M}\sum_{m=1}^{M} E_x\left[y_m(x) - f(x)^2\right]$$
$$E_{ENS} = E_x\left[\frac{1}{M}\sum_{m=1}^{M} y_m(x) - f(x)^2\right]$$

According to the Jensen's inequality we can state that $E_{ENS} \leq E_{AV}$ irrespective of E(y).