MACHINE LEARNING HOMEWORK SHEET-08

DEEP LEARNING

NAME: MAHALAKSHMI SABANAYAGAM TUM ID: ge73yuw

DATE: 16.12.2018

Problem 1:

Non linear activation functions in Neural Network:

The basis function for a neural network with k hidden layers is,

$$f(x, W) = \sigma_k (W_k^T \sigma_{k-1} (W_{k-1}^T \cdots \sigma_0 (W_0^T x)))$$

where $W_0, W_1, \ldots W_k$ are the weights in each layer

$\sigma_k, \sigma_{k-1} \ldots \sigma_0$ are the activation functions at each layer.

Say all the activation functions are linear then,

$$f(x, W) = (W_k^T W_{k-1}^T \cdots W_0^T)x$$

$$= (W')^T x$$

So, the basis function is just linear. Doing linear transformation at each layer is equivalent to having a single layer with linear transformation, putting k layers to no use. Also, while trying to learn using backpropogation gradient will not depend on the input in any layer and it will be a constant. This will lead to poor learning of the network.

Problem 2:

NN with a hidden layer having Sigmoid activation function.

To prove: An equivalent network exists which computes the same function but with $\tanh(x)$ as activation fn.

$$\text{Sigmoid } \sigma(x) = \frac{1}{1+e^{-x}}$$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

$$= \frac{\frac{1}{e^{-x}} - e^{-x}}{\frac{1}{e^{-x}} + e^{-x}}$$

$$= \frac{1 - e^{-2x}}{1+e^{-2x}} + 1 - 1$$

$$= \frac{1 - \cancel{e^{-2x}} + 1 + \cancel{e^{-2x}}}{1+e^{-2x}} - 1$$

$$= \frac{2}{1+e^{-2x}} - 1$$

$$\tanh(x) = 2\sigma(2x) - 1$$

$$\Rightarrow \tanh(x/2) = 2\sigma(x) - 1$$

$$\sigma(x) = \frac{1}{2}(\tanh(x/2) + 1)$$

$\therefore$ with the above relation, we see that if we apply $[\tan(x/2) + 1]\,1/2$ then it will be equivalent to the neural network using Sigmoid ~~action~~ activation function

Problem 3:

Derivative of $\tanh(x)$

$$\tanh(x) = \frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}$$

Applied quotient rule.

$$(\tanh(x))' = \frac{(e^{x}+e^{-x})(e^{x}+e^{-x}) - (e^{x}-e^{-x})(e^{x}-e^{-x})}{(e^{x}+e^{-x})^2}$$

$$= \frac{(e^{x}+e^{-x})^2 - (e^{x}-e^{-x})^2}{(e^{x}+e^{-x})^2}$$

$$= 1 - \left(\frac{e^{x} - e^{-x}}{e^{x} + e^{-x}}\right)^2$$

$$= 1 - (\tanh(x))^2$$

$$\therefore (\tanh(x))' = 1 - (\tanh(x))^2$$

Computing gradient in the backpropogation of neural network will be easier when $\tanh(x)$ is used as the activation function as the function $\tanh(x)$ would already be computed in the forward pass.

Problem 4:

$$y = \log \sum_{i=1}^{N} e^{x_i} \qquad y = a + \log \sum_{i=1}^{N} e^{x_i - a}$$

To prove : The above identity holds true.

$$y = a + \log \sum_{i=1}^{N} e^{x_i - a}$$

$$= a + \log \sum_{i=1}^{N} \frac{e^{x_i}}{e^{a}}$$

$$= a + \log \frac{1}{e^{a}} \sum_{i=1}^{N} e^{x_i} \quad (e^{a} \text{ independent of } i)$$

$$= a + \log e^{-a} + \log \sum_{i=1}^{N} e^{x_i} \quad (\log(ab) = \log a + \log b)$$

$$= \cancel{a} - \cancel{a} + \log \sum_{i=1}^{N} e^{x_i}$$

$$= \log \sum_{i=1}^{N} e^{x_i}$$

Thus, the identity is proved.

Problem 5 :

$$\frac{e^{x_i}}{\sum_{i=1}^{N} e^{x_i}} = \frac{e^{x_i - a}}{\sum_{i=1}^{N} e^{x_i - a}}$$

To prove : The above identity is true.

$$\frac{e^{x_i - a}}{\sum_{i=1}^{N} e^{x_i - a}} = \frac{\frac{e^{x_i}}{e^a}}{\sum_{i=1}^{N} \frac{e^{x_i}}{e^a}}$$

$$= \frac{\frac{e^{x_i}}{e^a}}{\frac{1}{e^a} \sum_{i=1}^{N} e^{x_i}}$$

$$= \frac{e^{x_i}}{\cancel{e^a}} \cdot \frac{\cancel{e^a}}{\sum_{i=1}^{N} e^{x_i}}$$

$$= \frac{e^{x_i}}{\sum_{i=1}^{N} e^{x_i}}$$

Thus the identity is proved.

Problem 6:

$$-\left(y \log \sigma(x) + (1-y) \log(1-\sigma(x))\right) \equiv \max(x,0) - xy + \log(1+e^{-abs(x)})$$

$$= -\left[\left(y \log (1+e^{-x})^{-1}\right) + \log\left(\frac{e^{-x}}{1+e^{-x}}\right) - y \log \frac{e^{-x}}{1+e^{-x}}\right]$$

$$= -\left[\cancel{-y\log(1+e^{-x})} + \log(e^{-x}) - \log(1+e^{-x}) - \underbrace{y \log e^{-x}}_{+yx} + \cancel{y \log(1+e^{-x})}\right]$$

$$= -\log(e^{-x}) + \log(1+e^{-x}) \mp xy$$

$\log(1+e^{-x})$:

To avoid overflow which will happen when

$1+e^{-x} \to \infty \Rightarrow e^{-x} \to \infty$

$\Rightarrow$ x is negative large no.

— We can avoid that by considering absolute value of x. $\therefore \log(1+e^{-abs(x)})$

$\log(e^{-x})$:

For negative values of x e^{-x} grows exponentially and log of that will be huge.

So, keeping the range of e^{-x} between $(0,1]$ will keep the value small and wont overflow.

$\therefore \log e^{-(\max(0,x))}$

Putting the above two in the original eq,

$$-\log e^{-(\max(0,x))} + \log(1+e^{-abs(x)}) - xy$$

$$= \max(0,x) + \log(1+e^{-abs(x)}) - xy.$$

Thus the equivalence is proved.