

Machine learning Homework- Deep Learning

Abinav Ravi Venkatakrishnan - 03694216 and Abhijeet Parida - 03679676

December 16, 2018

1 Activation Function

Problem 1:

The matrix operation $w^T + b$ is essentially a linear operation. When we stack linear operations over other linear operations we essentially get a linear function. It is impossible to approximate complex functions with just linear operations, therefore non-linearity is introduced to overcome this problem.

Problem 2:

The sigmoid activation function is

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

The tanh activation is

$$\begin{aligned} \tanh(x) &= \frac{e^{2x} - 1}{e^{2x} + 1} \\ \tanh\left(\frac{x}{2}\right) &= \frac{e^x - 1}{e^x + 1} \\ \tanh\left(\frac{x}{2}\right) &= \frac{1 - e^{-x}}{1 + e^{-x}} \\ \tanh\left(\frac{x}{2}\right) &= (1 - e^{-x})\sigma(x) \end{aligned}$$

Problem 3:

From the previous problem we know that,

$$\begin{aligned} \tanh(x) &= \frac{e^{2x} - 1}{e^{2x} + 1} \\ \tanh(x) &= \frac{e^x - e^{-x}}{e^x + e^{-x}} \end{aligned}$$

Taking derivative on both sides we get,

$$\begin{aligned} \frac{d}{dx}(\tanh(x)) &= \frac{d}{dx}\left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right) \\ &= (e^x + e^{-x})(e^x + e^{-x})^{-1} - (e^x + e^{-x})^{-2}(e^x - e^{-x})(e^x - e^{-x}) \\ &= 1 - \left(\frac{e^x - e^{-x}}{e^x + e^{-x}}\right)^2 \\ \frac{d}{dx}(\tanh(x)) &= 1 - (\tanh(x))^2 \end{aligned}$$

The advantage is that it is easy to compute the gradients during backpropagation.

2 Numerical Stability

Problem 4:

$$\begin{aligned} & a + \log(\sum_{i=1}^N e^{x_i - a}) \\ = & a + \log(\sum_{i=1}^N e^{x_i} e^{-a}) \\ = & a + \log(e^{-a} \sum_{i=1}^N e^{x_i}) \\ = & a + \log(e^{-a}) + \log(\sum_{i=1}^N e^{x_i}) \\ = & a - a + \log(\sum_{i=1}^N e^{x_i}) \\ = & \log(\sum_{i=1}^N e^{x_i}) \end{aligned}$$

Hence proved the equivalence

Problem 5:

$$\begin{aligned} & \frac{e^{x_i - a}}{\sum_{i=1}^N e^{x_i - a}} \\ = & \frac{e^{x_i} e^{-a}}{\sum_{i=1}^N (e^{x_i} e^{-a})} \end{aligned}$$

Hence proved the equivalence

Problem 6: