Machine Learning homework 8 solution
Deep Learning

Wiktor Jurasz - M.Nr. 03709419

December 16, 2018

1 Problem 1

With linear activation functions whole network reduces to linear transformation.

$$f(x, W) = W_k(W_{k-1}...(W_0x)) = W_kW_{k-1}...W_0x = W'x \tag{1}$$

Thus we would be limited only to linear functions.
With appropriate non linear activation functions, neural net is able to approximate any (differentiable) function. This property allows the net to find very complicated and complex patterns in data and create non linear mapping between input and output.

2 Problem 2

$$\sigma(x) = \frac{e^x}{e^x+1} = \frac{e^x - (\frac{1}{2}e^x + \frac{1}{2}) + (\frac{1}{2}e^x + \frac{1}{2})}{e^x+1} = \frac{\frac{1}{2}e^x - \frac{1}{2}}{e^x+1} + \frac{\frac{1}{2}(e^x+1)}{e^x+1} = \frac{1}{2}\frac{e^x-1}{e^x+1} + \frac{1}{2} = \frac{1}{2}tanh(\frac{1}{2}x) + \frac{1}{2} \tag{2}$$

From above we see that we can compute $\sigma(x)$ value using $tanh(x)$.

So instead of using $\sigma(w^Tx)$ as activation function, we can use:

$$g(w, x) = \frac{1}{2}tanh(\frac{1}{2}w^Tx) + \frac{1}{2} \tag{3}$$

To approximate the same function using a neural network.

3 Problem 3

$$tanh(x) = \frac{sinh(x)}{cosh(x)} \tag{4}$$

$$sinh(x) = \frac{e^x - e^{-x}}{2} \tag{5}$$

$$cosh(x) = \frac{e^x + e^{-x}}{2} \tag{6}$$

$$sinh'(x) = (\frac{e^x - e^{-x}}{2})' = \frac{1}{2}(e^x - e^{-x})' = \frac{1}{2}(e^x - \frac{-1}{e^{2x}} * e^x) = \frac{1}{2}(e^x + e^{-x}) = cosh(x) \tag{7}$$

$$cosh'(x) = (\frac{e^x + e^{-x}}{2})' = \frac{1}{2}(e^x + e^{-x})' = \frac{1}{2}(e^x + \frac{-1}{e^{2x}} * e^x) = \frac{1}{2}(e^x - e^{-x}) = sinh(x) \tag{8}$$

$$tanh'(x) = (\frac{sinh(x)}{cosh(x)})' = \frac{cosh(x)sinh'(x) - sinh(x)cosh'(x)}{cosh^2(x)} = \frac{cosh^2(x) - sinh^2(x)}{cosh^2(x)} = 1 - tanh^2(x) \tag{9}$$

Q.E.D

This property is useful, because we don't have to compute the derivative from scratch. If we already have the value of $tanh(x)$ it's enough to substitute it to the equation above to get derivative value.

4 Problem 4

$$a + log\sum_{i=1}^{N} e^{x_i - a} = a + log\sum_{i=1}^{N} e^{x_i}e^{-a} = a + log(e^{-a}\sum_{i=1}^{N} e^{x_i}) = a + log(e^{-a}) + log\sum_{i=1}^{N} e^{x_i} = a - a + log\sum_{i=1}^{N} e^{x_i} = log\sum_{i=1}^{N} e^{x_i} \tag{10}$$

Q.E.D

5 Problem 5

$$\frac{e^{x_i - a}}{\sum_{i=1}^{N} e^{x_i - a}} = \frac{e^{x_i}e^{-a}}{\sum_{i=1}^{N} e^{x_i}e^{-a}} = \frac{e^{x_i}e^{-a}}{e^{-a}\sum_{i=1}^{N} e^{x_i}} = \frac{e^{x_i}}{\sum_{i=1}^{N} e^{x_i}} \tag{11}$$

Q.E.D

6 Problem 6

$$\sigma(x) = \frac{e^x}{e^x + 1} \tag{12}$$

$$1 - \sigma(x) = 1 - \frac{e^x}{e^x + 1} = \frac{e^x + 1}{e^x + 1} - \frac{e^x}{e^x + 1} = \frac{1}{e^x + 1} \tag{13}$$

$$\begin{gathered} f(x) = -(ylog(\sigma(x)) + (1-y)log(1-\sigma(x))) = \\ -(ylog(\frac{e^x}{e^x+1}) + (1-y)log(\frac{1}{e^x+1})) = \\ -(y(log(e^x) - log(e^x+1)) + (1-y)(log(1) - log(e^x+1))) = \\ -(yx - ylog(e^x+1) + (1-y)(-log(e^x+1))) = \\ -(yx - ylog(e^x+1) - log(e^x+1) + ylog(e^x+1) = \\ -yx + log(e^x+1) = \\ -yx + log(e^x(1+e^{-x})) = \\ -yx + log(e^x) + log(1+e^{-x}) = \\ -yx + x + log(1+e^{-x}) \end{gathered} \tag{14}$$

Now we can also define f in this way.

$$f(x) = \begin{cases} -yx + log(e^x + 1) & x < 0 \\ -yx + x + log(1 + e^{-x}) & x \geq 0 \end{cases}$$

As it was shown above that both forms are equivalent.
From here we can also reduce this conditional function into one equation:

$$f(x) = -yx + max(x, 0) + log(1 + e^{-|x|}) \tag{15}$$

Q.E.D