Exercise 1

Hou, Jue
014695647
jue.hou@helsinki.fi

1. Activation functions

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

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

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

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

b) $\sigma'(x) = -(1+e^{-x})^{-2} \cdot e^{-x} \ \cdot -1 = (1+e^{-x})^{-2} \cdot e^{-x} = \frac{e^{-x}}{(1+e^{-x})^2}$

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

$\therefore\ \sigma'(x) = \sigma(x)\big(1 - \sigma(x)\big)$

c) $\tanh'(x) = 2\sigma'(2x)\ \cdot 2 = 4\sigma'(2x) = 4\sigma(2x)\big(1 - \sigma(2x)\big)$

d) ReLU

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

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

When it comes to $x > 0$, the result of the derivative of ReLU is always 1. When it comes to $x < 0$, the result of the derivative of ReLU is always 0. It will cause some problem if the derivative equals to zero. So, according to some online literature, it is usually use a very small value (e.g. 0.0000001) instead of 0 in practise. Or we can use some approximation function. For example, $f_k(x) = \frac{1}{2k}\log(1 + e^{-2kx})$, where the bigger k, the better approximation.

2. Forming neural networks