Exercise 1

Hou, Jue
014695647
jue.hou@helsinki.fi

1. Activation functions

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

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

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

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

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

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

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

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

d) ReLU

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

When it comes to $x > 0$, the derivative of ReLU is always 1. When it comes to $x < 0$, the derivative of ReLU is always 0. If the derivative equals to zero, then the corresponding gradients are zero as well. Some weights will therefore not be adjusted. It sometimes will cause problem. So, according to some online literature, people usually use a very small value (e.g. 0.0000001) instead of 0 in practise. Or we can use some approximation function. For example, $f_k(x) = \frac{1}{2k}\ln(1 + e^{-2kx})$, where the bigger k leads to a better approximation. Also, mathematically, the derivative of $f(0)$ does not exist. There are two ways to handle this problem. One is to assign a value for the derivative of $f(0)$. One common choice is simply 0. Another approach is to use the approximation functions. Just as what previous example has shown, they should be differentiable for every x.

2. Forming neural networks