Machine Learning Worksheet 7

Tomas Ladek, Michael Kratzer
3602673, 3612903
tom.ladek@tum.de, mkratzer@mytum.de

Problem 1

Let us rewrite the sigmoid activation function as

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

There exists a network that computes $\sigma(x)$ by scaling and offsetting the hyperbolic tangent function $\tanh(x)$, since

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

And therefore

$$\begin{aligned}\tanh(x) &= 2\sigma(2x) - 1 \\ \tanh(x) + 1 &= 2\sigma(2x) \\ \frac{1}{2}(\tanh(x)+1) &= \sigma(2x) \\ \frac{1}{2}(\tanh(\frac{z}{2})+1) &= \sigma(z) \qquad \textit{with } z = 2x\end{aligned}$$

Problem 2

$$\begin{aligned}\frac{d}{dx}\,\sigma(x) = \frac{d}{dx}\,\frac{1}{1+e^{-x}} &= \frac{d}{dx}\,(1+e^{-x})^{-1} = -(1+e^{-x})^{-2}\cdot(-e^{-x}) \\ &= \frac{e^{-x}}{(1+e^{-x})^2} = e^{-x}\Big(\frac{1}{1+e^{-x}}\Big)^2 = e^{-x}\sigma^2(x)\end{aligned}$$

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