Machine Learning Worksheet 07

Shang-Hsin Yu – 03681048 – shanghsin.yu@tum.de

Problem 1

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

The *tanh* function is just a scaled version of the Σ function.
With a different value in the learning result they can produce the same result.

Problem 2

For σ:

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

For *tanh*:

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

Problem 3

Trying to solve the MLE we need to minimize the negative log likelihood

$$= \frac{1}{2}\beta \sum_{k=1}^{K} (y(x_k, w) - z_k)^T (y(x_k, w) - z_k)$$

which is basically the same as the the sum-of-square function

$$\frac{1}{2} \sum_{k=1}^{K} (y(x_k, w) - z_k)^2$$

Problem 4

Similar to the last exercise, we take the negative log likelihood of the distribution and drop unimportant terms with respect to w and we get

$$\frac{1}{\beta}\sum_{n=1}^{N} \mid z_n - y(x_n, w) \mid$$

which is basically the same as minimizing

$$\sum_{n=1}^{N} \mid z_n - y(x_n, w) \mid$$

Problem 5

It is very strange that for the sin function the more hidden nodes doesn't yield better result. which don't make sense to me...

Problem 6

It is to be used as the biased term.

Problem 7

There are not enough training data for the neural network to learn sufficiently?