My solutions to

Deep Learning: Foundations and Concepts

Dario Miro Konopatzki

4 Single-Layer Networks: Regression

4.1

$$
\begin{aligned}
0 &\stackrel{!}{=} \frac{\partial}{\partial w_i} E(w) \\
&= \frac{\partial}{\partial w_i} \frac{1}{2} \sum_{n=1}^{N} \left(\sum_{j=0}^{M} w_j x_n^j - t_n \right)^2 \\
&= \frac{1}{2} \sum_{n=1}^{N} 2 \left(\sum_{j=0}^{M} w_j x_n^j - t_n \right) x_n^i \\
&= \sum_{n=1}^{N} \sum_{j=0}^{M} w_j x_n^{j+i} - \sum_{n=1}^{N} t_n x_n^i \\
\iff \quad & \sum_{j=0}^{M} \sum_{n=1}^{N} x_n^{i+j} w_j = \sum_{n=1}^{N} x_n^i t_n \\
\iff \quad & \sum_{j=0}^{M} A_{ij} w_j = T_i
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2}{\partial w_k \partial w_i} E(w) &= \frac{\partial}{\partial w_k} \left(\sum_{n=1}^{N} \sum_{j=0}^{M} w_j x_n^{j+i} - \sum_{n=1}^{N} t_n x_n^i \right) \\
&= \sum_{n=1}^{N} x_n^{k+i}
\end{aligned}
$$

I.e. $\mathrm{H}_E = X^\top X$. For $v \neq 0$, $v^\top X^\top X v = (Xv)^\top X v = \|Xv\|_2^2 \geq 0$, i.e. H_E positive semidefinite, and hence positive definite if and only if it is nonsingular, in which case the solution to $Aw = T$ minimizes $E(w)$.

4.3

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

$$
\begin{aligned}
\implies 2\sigma(2a) &= \tanh(a) + 1 \\
\Leftrightarrow \sigma(2a) &= \frac{1}{2}\tanh(a) + \frac{1}{2} \\
\Leftrightarrow \sigma(a) &= \frac{1}{2}\tanh\left(\frac{a}{2}\right) + \frac{1}{2}
\end{aligned}
$$

$$
\begin{aligned}
\implies y(x,w) &= w_0 + \sum_{j=1}^{M} w_j \sigma\left(\frac{x-\mu_j}{s}\right) \\
&= w_0 + \sum_{j=1}^{M} w_j \left(\frac{1}{2}\tanh\left(\frac{x-\mu_j}{2s}\right) + \frac{1}{2}\right) \\
&= \underbrace{w_0 + \sum_{j=1}^{M} \frac{w_j}{2}}_{:=u_0} + \sum_{j=1}^{M} \underbrace{\frac{w_j}{2}}_{:=u_j} \tanh\left(\frac{x-\mu_j}{2s}\right)
\end{aligned}
$$