My solutions to

Deep Learning: Foundations and Concepts

Dario Miro Konopatzki

5 Single-Layer Networks: Classification

5.1

For all $1 \leq k \leq K$:

$$\begin{aligned}
\mathbb{E}\left[T_k | X = x\right] &= \sum_{t \in \{0,1\}} t\, p_{T_k|X=x}(t) \\
&= 0 \cdot p_{T_k|X=x}(0) + 1 \cdot p_{T_k|X=x}(1) \\
&= p_{T_k|X=x}(1) \\
&= \begin{cases} 1 & \text{if class of } x \text{ is } c_k \\ 0 & \text{otherwise} \end{cases} \\
&= p_{C|X=x}(c_k)
\end{aligned}$$

Hence $\mathbb{E}\left[T | X = x\right] = \left(p_{C|X=x}(c_k)\right)_{1 \leq k \leq K}$.

5.2

Assume the convex hulls intersect, i.e. t.e. (α_n) and (β_m) s.t.

$$z := \sum_n \alpha_n x_n = \sum_m \beta_m y_m$$

Now assume the sets are linearly separable, i.e. t.e. $\hat{w}, w_0$ s.t. $\hat{w}^\top x_n + w_0 > 0$ for all x_n and $\hat{w}^\top y_m + w_0 < 0$ for all y_m. Then

$$\hat{w}^\top z + w_0 = \hat{w}^\top \left(\sum_n \alpha_n x_n \right) + w_0$$

$$
\begin{aligned}
&= \sum_n \alpha_n \hat{w}^\top x_n + w_0 \\
&> \sum_n \alpha_n (-w_0) + w_0 \qquad \alpha_n \geq 0 \text{ f.a. } n \\
&= -w_0 \sum_n \alpha_n + w_0 \\
&= -w_0 \cdot 1 + w_0 \qquad \sum_n \alpha_n = 1 \\
&= 0
\end{aligned}
$$

but also, $\hat{w}^\top z + w_0 < 0$ by the analogous argument via $z = \sum_m \beta_m y_m$, which is a contradiction.

The converse statement, that if the two sets of points are not linearly separable, then their convex hulls intersect, follows immediately from the contrapositive of the separating hyperplane theorem.

5.18

$$
\begin{aligned}
\frac{\mathrm{d}\sigma}{\mathrm{d}a} &= \frac{\mathrm{d}}{\mathrm{d}a} \frac{1}{1+e^{-a}} \\
&= \frac{0 \cdot (1+e^{-a}) - 1 \cdot (-1) e^{-a}}{(1+e^{-a})^2} \\
&= \frac{e^{-a}}{(1+e^{-a})^2} \\
&= \frac{1}{1+e^{-a}} \frac{e^{-a}}{1+e^{-a}} \\
&= \sigma(a) \frac{1+e^{-a}-1}{1+e^{-a}} \\
&= \sigma(a)\,(1-\sigma(a))
\end{aligned}
$$