BLOCK II EXERCISES:

LINEAR CLASSIFICATION

[SOLUTIONS]

$\boxed{\text{Ex. 1}}$

i) $1 - \sigma(a) = 1 - \dfrac{1}{1-e^{-a}} = \dfrac{1-e^{-a}-1}{1-e^{-a}} = \dfrac{e^{-a}}{1-e^{-a}} \times\left(\dfrac{1}{e^{-a}}\right) = \dfrac{1}{e^{a}-1} = \sigma(-a)$

ii) Let us define $y = \sigma(a)$ and $a = \sigma^{-1}(y)$

$$y = \frac{1}{1+e^{-a}}$$

$$1+e^{-a} = \frac{1}{y}$$

$$e^{-a} = \frac{1}{y} - 1$$

$$-a = \ln\left(\frac{1}{y} - 1\right)$$

$$a = -\ln\left(\frac{1-y}{y}\right)$$

$$a = \ln\left(\frac{y}{1-y}\right) = \sigma^{-1}(y)$$

$\boxed{\text{Ex. 7}}$

We know the ML solution for the logistic regression model is

$$w^* = \underset{w}{\operatorname{argmin}} \frac{1}{n} \sum_{i=1}^{n} \log\left(1 + e^{\left(-y_i w^T \phi(x_i)\right)}\right)$$

We know if the data is linearly separable, any decision boundary that separates both classes fulfills:

$$w^T \phi(x_i) \begin{cases} \geq 0 & \text{if } y_i = 1 \\ < 0 & \text{otherwise.} \end{cases}$$

An equivalent way to write the ML solution is

$$P(Y = y_i \mid X = x_i, w) = \sigma(w^T \phi(x_i))$$

$$w^* = \underset{w}{\operatorname{argmax}} \prod_i P(Y = y_i \mid X = x_i, w)$$

which is maximized when

$$\hat{y}_i = \sigma(w^T \phi(x_i)) = \begin{cases} 1 & \text{if } y_i = 1 \\ 0 & \text{otherwise} \end{cases}$$

That is, it is maximized when the sigmoid saturates, which happens when $w^T \phi(x_i) \to \pm \infty$, i.e. when the magnitude of w goes to ∞

Ex. 3

One-hot encoding!

$\varphi_n = \phi_n(x_n)$

The likelihood function is

$$p(\{\varphi_n, y_n\} | \{\pi_k\}) = \prod_n p(\varphi_n, y_n | \{\pi_k\})$$

identifies the class of n

$$= \prod_n p(\varphi_n | y_n, \{\pi_k\}) \, p(y_n | \{\pi_k\})$$

$$= \prod_n \prod_k \left[p(\varphi_n | C_k) \, p(C_k) \right]^{y_{nk}}$$

$y_{nj} \in \{0, 1\}$

$y_n = [y_{n1}, \ldots, y_{nk}]$

$\sum_{j=1}^{k} y_{nj} = 1$

If we take the log

$$\log p(\{\varphi_n, y_n\} | \{\pi_k\}) = \sum_n \sum_k y_{nk} \left[\log p(\varphi_n | C_k) + \log p(C_k) \right]$$

$$= \sum_n \sum_k y_{nk} \left[\log p(\varphi_n | C_k) + \log \pi_k \right]$$

Then the ML solution for π_k is

$$\pi_k^* = \operatorname*{argmax}_{\pi_k} \log(\{\varphi_n, y_n\} | \{\pi_k\})$$

$$\text{s.t.} \quad \sum_k \pi_k = 1$$

such that

To solve this constraint optimization problem, we rely on Lagrange multipliers.

$$\pi_k^* = \operatorname*{argmax}_{\pi_k} \Big[\underbrace{\log p(\{\varphi_n, y_n\} | \{\pi_k\}) + \lambda \Big(\sum_k \pi_k - 1 \Big)}_{L(\{\varphi_n, y_n\}, \{\pi_k\}, \lambda)} \Big]$$

We set the derivative to zero

$$\frac{\partial L(\pi_k)}{\partial \pi_k} = \sum_n \frac{y_{nk}}{\pi_k} + \lambda = 0 \rightarrow \sum_n y_{nk} = -\pi_k \lambda$$

$$\rightarrow N_k = -\pi_k \lambda \rightarrow \lambda = \frac{-N_k}{\pi_k}$$

↑ number of training samples with class k

Now we set $\frac{\partial L(\lambda)}{\partial \lambda} = 0$

$$\frac{\partial L(\lambda)}{\partial \lambda} = 0 \longrightarrow \sum_k \pi_k = 1 \longrightarrow \sum_k -\frac{N_k}{\lambda} = 1 \rightarrow$$

$$\lambda = -N$$

Finally, we have that $\boxed{\pi_k^* = \frac{N_k}{N}}$