Ex1:

(a) Formula of the sigmoid function:

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

- The derivative of the sigmoid function:

$$\sigma'(a) = \left(\frac{1}{1+e^{-a}}\right)'$$

$$= -\frac{1}{(1+e^{-a})^2} (1+e^{-a})'$$

$$= \frac{e^{-a}}{(1+e^{-a})^2}$$

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

$$= \sigma(a)(1-\sigma(a))$$

(b) Formula of the loss function in logistic regression

$$L = -\sum_{n=1}^{N} (t_n \log y_n + (1-t_n)\log(1-y_n))$$

where
$$\begin{cases} y_n = \sigma(a_n) \\ a_n = w^T \phi_n \\ \phi_n = \phi(x_n) \\ t_n \in \{0; 1\} \end{cases}$$

$\Rightarrow$ Cross_entropy Loss

(c) Calculate the gradient vector for loss function in logistic regression

$$\nabla L = \frac{\partial L}{\partial w} = \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial a} \times \frac{\partial a}{\partial w}$$

• $\frac{\partial y}{\partial a} = \frac{\partial \sigma(a)}{\partial a} = \sigma(a)(1 - \sigma(a)) = y(1-y)$

• $\frac{\partial a}{\partial w} = \frac{\partial (w^T \phi)}{\partial w}$

$$w = \begin{bmatrix} w_0 \\ w_1 \\ \vdots \\ w_n \end{bmatrix} ; \quad \phi = \begin{bmatrix} 1 \\ \phi_1 \\ \vdots \\ \phi_n \end{bmatrix}$$

$$\Rightarrow w^T \phi = w_0 + w_1 \phi_1 + \dots + w_n \phi_n$$

$$\Rightarrow \frac{\partial a}{\partial w_t} = \phi_t \quad (t \in [1, n])$$

$$\Rightarrow \frac{\partial a}{\partial w} = \begin{bmatrix} \frac{\partial a}{\partial w_0} \\ \frac{\partial a}{\partial w_1} \\ \vdots \\ \frac{\partial a}{\partial w_n} \end{bmatrix} = \begin{bmatrix} 1 \\ \phi_1 \\ \vdots \\ \phi_n \end{bmatrix} = \phi$$

• $\nabla L = \frac{\partial L}{\partial y} \times \frac{\partial y}{\partial a} \times \frac{\partial a}{\partial w}$

$$= -\sum_{n=1}^{N} \frac{\partial L}{\partial y_n} \left(t_n \log y_n + (1 - t_n) \log (1 - y_n) \right) \times \frac{\partial y}{\partial a_n} \times \frac{\partial a}{\partial w_n}$$

$$= -\sum_{n=1}^{N} \left(t_n \frac{1}{y_n} + (1 - t_n) \frac{-1}{1 - y_n} \right) \times \frac{\partial y}{\partial a_n} \times \phi_n$$

$$= -\sum_{n=1}^{N} \left(t_n \frac{y_n'}{y_n} - (1 - t_n) \frac{y_n'}{1 - y_n} \right) \times \phi_n$$

$$= -\sum_{n=1}^{N} \left(t_n \frac{y_n(1 - y_n)}{y_n} - (1 - t_n) \frac{y_n(1 - y_n)}{1 - y_n} \right) \times \phi_n$$

KLONG

$$= -\sum_{n=1}^{N} \left(t_n (1-y_n) - (1-t_n) y_n \right) \phi_n$$

$$= -\sum_{n=1}^{N} \left(t_n - t_n y_n - y_n + t_n y_n \right) \times \phi_n$$

$$= -\sum_{n=1}^{N} (t_n - y_n) \phi_n$$

$$= \sum_{n=1}^{N} (y_n - t_n) \phi_n$$

Hence, $\nabla L = \sum_{n=1}^{N} (y_n - t_n) \phi_n$

$$= [1 \ \phi_1 \ldots \phi_n] \begin{bmatrix} y_1 - t_1 \\ y_2 - t_2 \\ \vdots \quad \vdots \\ y_n - t_n \end{bmatrix}$$

$$\frac{\partial L}{\partial w} = \phi^T (\hat{y} - t)$$