Q5. $p(C_1|\vec{x}) = \frac{1}{1+e^{-a(\vec{x})}} = g(a(\vec{x}))$

$$a(x) = \ln \frac{p(\vec{x}|C_1)p(C_1)}{p(\vec{x}|C_2)p(C_2)}$$

Assume Linear decision boundary. $a(\vec{x}) = \vec{w}^T\vec{x}$.

Goal: Select $\vec{w}^*$ to maximize $P(Y|X, \vec{w})$

$$\underbrace{\vec{w}^* = \begin{bmatrix} b \\ w_1 \\ w_2 \end{bmatrix}}_{\uparrow \text{ parameter to learn}} \quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

sigmoid function

$$g(a(\vec{x})) = \frac{1}{1+e^{-a(\vec{x})}}$$

where $a(\vec{x}) = \vec{w}^T \cdot \vec{x} = b + w_1x_1 + w_2x_2$

where $\vec{w}^T = \begin{bmatrix} b \\ w_1 \\ w_2 \end{bmatrix}$, $\vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

$$p(Y, X|\vec{w}) = p(Y|X, \vec{w}) \cdot p(X|\vec{w})$$

$$\propto p(Y|X, \vec{w}) = \prod_{i=1}^{N} P(y_i|\vec{x}_i, \vec{w}$$

$$\underline{P(\vec{y}|X, \vec{w})} = \prod_{i=1}^{N} p(y_i|\vec{x}_i, \vec{w})$$

assume training data points

are independently distributed

$$P_i \equiv P(y = C_i|\vec{x}_i, \vec{w})$$

$$\prod_{i=1}^{N} P(y_i|\vec{x}_i, \vec{w}) = \Pi_{i:y_i=1} P_i \; \Pi_{i:y_i=0}(1-P_i) = \prod_{i=1}^{N} P_i^{y_i}(1-P_i)^{(1-y_i)}$$

$$\vec{w}_{ML} = \underset{\vec{w}}{\operatorname{argmax}}\, P(Y|X, \vec{w})$$

$$= \underset{\vec{w}}{\operatorname{argmin}} -\log P(Y|X, \vec{w})$$

Objective function

$$P_i = g(\vec{w}^T x_i)$$

$$E(\vec{w}) = -\log \prod_{i=1}^{N} P_i^{y_i}(1-P_i)^{(1-y_i)}$$

$$\frac{d}{d\alpha} g(\alpha) = g(\alpha)(1-g(\alpha))$$

$$\begin{cases} \frac{d}{d\vec{w}} \log(g(a)) = (1-g(a))\vec{x_i} \\ \frac{d}{d\vec{w}} \log(1-g(a)) = -g(a)\vec{x_i} \end{cases}$$

$$\Rightarrow \nabla E(\vec{w}) = -\sum_{i=1}^{N} [y_i(1-g(a))\vec{x_i} - (1-y_i)g(a)\vec{x_i}]$$

$$= -\sum_{i=1}^{N} y_i\vec{x_i} - \cancel{y_i g(a)}\vec{x_i} - g(a)\vec{x_i} + \cancel{y_i g(a)}x_i)$$

$$= -\sum_{i=1}^{N} y_i\vec{x_i} - g(a)\vec{x_i}$$

$$= -\sum_{i=1}^{N} (y_i - g(a))\vec{x_i}$$