1.1

$$J(w, w_0) = (y^T - w^T X^T - w_0 1^T)(y - Xw - w_0 1) + \lambda w^T w$$

$$= y^T y - y^T X w - w_0 y^T 1 - w^T X^T y + w^T X^T X w + w_0 w^T X^T 1$$

$$- w_0 1^T y + w_0 1^T X w + n w_0^2 + \lambda w^T w$$

$$= y^T y - 2 y^T X w - 2 w_0 y^T 1 + w^T X^T X w + 2 w_0 w^T X^T 1 + n w_0^2 + \lambda w^T w$$

since $\bar{x} = \frac{1}{n} \sum_{i=1}^{n} x_i = 0$. so $\sum_{i=1}^{n} x_i = 0$. $\Rightarrow X^T 1 = 0$

$$J(w, w_0) = y^T y - 2 y^T X w - 2 w_0 y^T 1 + w^T X^T X w + n w_0^2 + \lambda w^T w$$

$$\frac{\partial}{\partial w_0} J(w, w_0) = -2 y^T 1 + 2 n w_0 = 0$$

$$w_0 = \frac{1}{n} y^T 1 = \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y}$$

$$\frac{\partial}{\partial w} J(w, w_0) = -2 y^T X + 2 w^T X^T X + 2 \lambda w^T = 0$$

$$w^T (X^T X + \lambda I) = y^T X$$

$$(X^T X + \lambda I) w = X^T y$$

$$w = (X^T X + \lambda I)^{-1} X^T y$$

1.2

(i) $\phi(0) = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}$ $\phi(1) = \begin{bmatrix} 1 \\ \sqrt{2} \\ 1 \end{bmatrix}$

$$w^T(\phi(0) - \phi(1)) = 0$$

$$w^T \begin{bmatrix} 0 \\ -\sqrt{2} \\ -1 \end{bmatrix} = 0$$

the vector is orthogonal to $\begin{bmatrix} 0 \\ -\sqrt{2} \\ -1 \end{bmatrix}$ and Parallel to w

it can be $\begin{bmatrix} 1 \\ 1 \\ -\sqrt{2} \end{bmatrix}$

(ii)

$$d = \sqrt{(1-1)^2 + (\sqrt{2}-0)^2 + (0-1)^2} = \sqrt{3}$$

$$\text{Margin} = \frac{d}{2} = \frac{\sqrt{3}}{2}$$

(iii)

$$-1(w_1 + 0 + 0 + w_0) \geq 1$$

$$1(w_1 + \sqrt{2}w_2 + w_3) \geq 1$$

$$\begin{cases} w_0 + w_1 \leq -1 \\ w_1 + \sqrt{2}w_2 + w_3 + w_0 \geq 1 \end{cases} \Rightarrow \begin{cases} w_0 + w_1 = -1 \\ w_1 + \sqrt{2}w_2 + w_3 + w_0 = 1 \end{cases} \Rightarrow \sqrt{2}w_2 + w_3 = 2$$

$$\min \|w\|^2 = w_1^2 + w_2^2 + w_3^2 = w_1^2 + w_2^2 + (2 - \sqrt{2}w_2)^2$$

$$= w_1^2 + 3w_2^2 - 4\sqrt{2}w_2 + 4$$

$$\begin{cases} \frac{\partial}{\partial w_1}\|w\|^2 = 2w_1 = 0 \\ \frac{\partial}{\partial w_2}\|w\|^2 = 6w_2 - 4\sqrt{2} = 0 \\ \sqrt{2}w_2 + w_3 = 2 \end{cases} \Rightarrow \begin{cases} w_1 = 0 \\ w_2 = \frac{2\sqrt{2}}{3} \\ w_3 = \frac{2}{3} \end{cases}$$

$$w = \begin{bmatrix} 0 \\ \frac{2\sqrt{2}}{3} \\ \frac{2}{3} \end{bmatrix} \qquad \frac{1}{\|w\|} = \frac{1}{\sqrt{0^2 + \left(\frac{2\sqrt{2}}{3}\right)^2 + \left(\frac{2}{3}\right)^2}} = \frac{1}{\sqrt{\frac{4}{3}}} = \frac{\sqrt{3}}{2}$$

(iv) $w_0 + w_1 = -1$

$w_0 = -1$

(v)

$$f(x) = -1 + \left[0, \frac{2\sqrt{2}}{3}, \frac{2}{3}\right] \begin{bmatrix} 1 \\ \sqrt{2}x \\ x^2 \end{bmatrix} = \frac{2}{3}x^2 + \frac{4}{3}x - 1$$

2.1

$$L = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log(\hat{y}_{ic}) = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\log\left(\frac{\exp(w_c^T x)}{\sum_{j=1}^{C}\exp(w_j^T x)}\right)$$

$$\frac{\partial L}{\partial w} = -\frac{1}{N}\sum_{i=1}^{N}\sum_{c=1}^{C} y_{ic}\frac{1}{\hat{y}_{ic}}\frac{\partial \hat{y}_{ic}}{\partial w}$$

$$\frac{\partial \hat{y}_{ic}}{\partial w_c}\begin{cases}\hat{y}_{ic}(1-\hat{y}_{ic})x_i, & c=k.\\ -\hat{y}_{ic}\hat{y}_{ik}x_i, & c\neq k.\end{cases}$$

$$\frac{\partial L}{\partial w} = -\frac{1}{N}\sum_{i=1}^{N}\left[y_{ic}\frac{1}{\hat{y}_{ic}}\hat{y}_{ic}(1-\hat{y}_{ic})x_i + \sum_{c\neq k} y_{ic}\frac{1}{\hat{y}_{ic}}\cdot(-\hat{y}_{ic}\hat{y}_{ik}x_i)\right]$$

$$= \frac{1}{N}\sum_{i=1}^{N}[\hat{y}_{ic} - y_{ic}]x_i$$