2.1

(a)

$X_{K\times N}$

$W_{1\times K}$

$y_{1\times N}$

$$f(w) \triangleq \log p(y^{(1)}, \dots, y^{(N)} \mid \vec{x}^{(1)}, \dots \vec{x}^{(N)}, \vec{w})$$

$$= \sum_{i=1}^{N} \log p(y^{(i)} \mid x^{(i)}, \vec{w}) = \sum_{i=1}^{N} \begin{cases} -\log p_i(w) & \text{if } y^{(i)}=1 \\ -\log(1-p_i(w)) & \text{if } y^{(i)}=0 \end{cases} \quad \text{where } p_i(w) = \frac{1}{1+\exp(-\vec{w}^T x)}$$

$$= \sum_{i=1}^{N} y^{(i)} \log p_i(w) + (1-y^{(i)}) \cdot \log(1-p_i(w))$$

(b)

$$\frac{\partial f_i(w)}{\partial w_j} = \underbrace{\boxed{y^{(i)}(\log p_i(w))}}_{\langle \text{Left side} \rangle} + \underbrace{\boxed{(1-y^{(i)})(\log(1-p_i(w)))}}_{\langle \text{Right side} \rangle}$$

<Left>

$$y^{(i)} \cdot \frac{1}{p_i(w)} \cdot \frac{d}{dw_j}\left(\frac{1}{1+\exp(-w^T x^{(i)})}\right) = y^{(i)} \cdot \frac{1}{p_i(w)} \cdot -\left(1+\exp(-w^T x^{(i)})\right)^{-2} \cdot \exp(-w^T x^{(i)}) \cdot (-x_j^{(i)})$$

$$= \frac{y^{(i)} \cdot \exp(-w^T x^{(i)})}{1+\exp(-w^T x^{(i)})} x_j^{(i)} = y^{(i)} \cdot (1-p(w))\, x_j^{(i)}$$

<Right> 令 $1-p_i(w) = \lambda$

$$\frac{d}{dw_j}(1-y^{(i)})(\log(\lambda)) \frac{d\lambda}{dp} = (1-y^{(i)})(1-\lambda) \cdot x_j^{(i)} \cdot (-1) = (1-y^{(i)})(p_i(w)) \cdot -x_j^{(i)}$$

$$\frac{\partial f_i(w)}{\partial w_j} = \text{left} + \text{right} = (y^{(i)} - p_i(w))\, x_j^{(i)}$$

$$\frac{\partial f(w)}{\partial w_j} = \frac{\partial f_i(w)}{\partial w_j} = (y^{(i)} - p(w))\, x_j^{(i)}$$

(c)

$$w_j = w_j + \eta \frac{\partial f(w)}{\partial w_j}$$

$\uparrow$

step size