Ex 2-1

(a) $z = Wx = \begin{pmatrix} w_{11} & \cdots & w_{1d} \\ \vdots & \ddots & \vdots \\ w_{n1} & \cdots & w_{nd} \end{pmatrix} \begin{pmatrix} x_1 \\ \vdots \\ x_d \end{pmatrix} = \begin{pmatrix} \sum_{k=1}^{d} w_{1k} x_k \\ \vdots \\ \sum_{k=1}^{d} w_{nk} x_k \end{pmatrix}$

$J_z = \frac{\partial z}{\partial x} \Rightarrow (J_z)_{ij} = \frac{\partial z_i}{\partial x_j} = \frac{\partial}{\partial x_j}\left(\sum_{k=1}^{d} w_{ik} x_k\right) = w_{ij}$.

Thus: $J_z = W$

(b) $z = x^T W^T = (Wx)^T = (z_1, \cdots, z_n)$

$J_z = \frac{\partial z}{\partial x} = \begin{pmatrix} \frac{\partial z_1}{\partial x_1}, & \cdots, & \frac{\partial z_n}{\partial x_1} \\ \vdots & & \\ \frac{\partial z_1}{\partial x_d}, & \cdots, & \frac{\partial z_n}{\partial x_d} \end{pmatrix}_{d \times n} \Rightarrow \left(\frac{\partial z}{\partial x}\right)_{ij} = \frac{\partial z_j}{\partial x_i} = w_{ji}$.

Thus: $J_z = W^T$

(c) $z_i = f(x_i) \quad \frac{\partial z_i}{\partial x_j} = \frac{\partial f(x_i)}{\partial x_j} = \begin{cases} \frac{\partial f(x_i)}{\partial x_i} & \text{if } i=j \\ 0 & \text{else} \end{cases}$.

Thus: $J_z = \begin{pmatrix} \frac{\partial f(x_1)}{\partial x_1} & & 0 \\ & \ddots & \\ 0 & & \frac{\partial f(x_d)}{\partial x_d} \end{pmatrix} = \text{diag}(f'(x))$

(d) $L = L(z) \quad z = Wx$. $\quad \frac{\partial L}{\partial W} = \begin{pmatrix} \frac{\partial L}{\partial w_{11}} & \cdots & \frac{\partial L}{\partial w_{1d}} \\ \vdots & \ddots & \vdots \\ \frac{\partial L}{\partial w_{n1}} & \cdots & \frac{\partial L}{\partial w_{nd}} \end{pmatrix}$

$\Rightarrow \left(\frac{\partial L}{\partial W}\right)_{ij} = \frac{\partial L}{\partial w_{ij}} = \sum_{k=1}^{n} \frac{\partial L}{\partial z_k} \frac{\partial z_k}{\partial w_{ij}}$ (chain rule)

$\frac{\partial L}{\partial z_k} = (\nabla L)_k \quad \frac{\partial z_k}{\partial w_{ij}} = \frac{\partial}{\partial w_{ij}}\left(\sum_{l=1}^{n} w_{kl} x_l\right) = \begin{cases} x_j & \text{if } i=k,\ l=j \\ 0 & \text{else} \end{cases}$.

Thus: $\left(\frac{\partial L}{\partial W}\right)_{ij} = (\nabla L)_i \cdot x_j \Rightarrow \frac{\partial L}{\partial W} = \nabla L \cdot x^T$

Ex 2-2

(a) $$\hat{y}_i = \frac{e^{z_i}}{\sum_{k=1}^{N} e^{z_k}}$$

$$\frac{\partial \hat{y}_i}{\partial z_j} = \frac{\frac{\partial e^{z_i}}{\partial z_j} \cdot \sum_{k=1}^{N} e^{z_k} - e^{z_i} \cdot \frac{\partial \sum_{k=1}^{N} e^{z_k}}{\partial z_j}}{\left(\sum_{k=1}^{N} e^{z_k}\right)^2}$$

If $i=j$:

$$\frac{\partial \hat{y}_i}{\partial z_j} = \frac{e^{z_i} \cdot \sum_{k=1}^{N} e^{z_k} - e^{z_i} \cdot e^{z_i}}{\left(\sum_{k=1}^{N} e^{z_k}\right)^2} = \frac{e^{z_i}}{\sum_{k=1}^{N} e^{z_k}} \cdot \left(1 - \frac{e^{z_i}}{\sum_{k=1}^{N} e^{z_k}}\right)$$

$$= \hat{y}_i (1 - \hat{y}_i)$$

If $i \neq j$:

$$\frac{\partial \hat{y}_i}{\partial z_j} = \frac{0 - e^{z_i} \cdot e^{z_j}}{\left(\sum_{k=1}^{N} e^{z_k}\right) \cdot \left(\sum_{k=1}^{N} e^{z_k}\right)} = -\hat{y}_i \hat{y}_j$$

(b) $$L = -\sum_{k=1}^{N} y_k \log \hat{y}_k$$

$$\frac{\partial L}{\partial z_i} = \sum_i \left(\frac{\partial L}{\partial \hat{y}_i} \frac{\partial \hat{y}_i}{\partial z_j} + \overbrace{\frac{\partial L}{\partial y_i} \cdot \frac{\partial y_i}{\partial z_j}}^{0} \right) = \sum_i \left(\frac{\partial L}{\partial \hat{y}_i} \cdot \frac{\partial \hat{y}_i}{\partial z_j} \right)$$

$$= -\frac{y_j}{\hat{y}_j} \cdot \hat{y}_j (1 - \hat{y}_j) + \sum_{i,\, i \neq j} \left(-\frac{y_i}{\hat{y}_i} \right) \cdot \left(-\hat{y}_i \cdot \hat{y}_j \right)$$

$$= -y_j + y_j \hat{y}_j + \sum_{i,\, i \neq j} y_i \cdot \hat{y}_j$$

$$= -y_j + \hat{y}_j \sum_i y_i$$

$$= \hat{y}_j - y_j \qquad \left(\text{due to } \sum_i y_i = 1\right)$$

Ex 2-3

$$L(y,\hat{y}) = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2 = \frac{1}{n}(\hat{y}-y)^T(\hat{y}-y)$$

$$= \frac{1}{n}\left(\hat{y}^T\hat{y} - 2\hat{y}^T y + y^T y\right)$$

$$\hat{y} = Xw \Rightarrow L(y,\hat{y}) = \frac{1}{n}\left(w^T X^T X w - 2 w^T X^T y + y^T y\right)$$

Thus :

$$\frac{\partial L}{\partial w} = \frac{1}{n}\left(2X^T X w - 2X^T y + 0\right) = 0$$

$$\Rightarrow \quad 2X^T X w - 2X^T y = 0$$

$$\Rightarrow \quad X^T X w = X^T y \Rightarrow w = (X^T X)^{-1} X^T y$$

$$(\text{if } |X^T X| \neq 0)$$

det