ML $\quad \underline{\theta} = \underbrace{(X^T X)^{-1}}_{D \times D} \underbrace{X^T}_{D \times N} \underbrace{\underline{y}}_{N \times 1}$ $\quad X : N \times D$

$D^3 \quad y : N \times 1$

$$\mu = \underbrace{X^T}_{D \times N} \underbrace{(X X^T)^{-1}}_{N \times N} \underbrace{\underline{y}}_{N \times 1} \qquad N^3$$

$$\theta = (X^T X)^{-1} X^T y$$

$$= (X^T X)^{-1} X^T \cdot I_N y$$

$$= (X^T X)^{-1} X^T \underbrace{(X X^T) \cdot}(X X^T)^{-1} y$$

$$= \underbrace{(X^T X)^{-1} \cdot (X^T X)}_{I_D} \cdot X^T (X X^T)^{-1} y$$

$$\boxed{\theta = X^T (X X^T)^{-1} y}$$

$$(X^T X + \lambda I_D)^{-1} X^T y = X^T (X X^T + \lambda I_N)^{-1} y$$

$$L(\theta) = \frac{1}{N} \sum_n L_n \qquad \frac{dL}{d\theta} = \frac{1}{N} \sum_n \frac{dL_n}{d\theta}$$

$$L_n = -y_n \log g_\theta(x_n) - (1 - y_n) \log(1 - g_\theta(x_n))$$

$$\frac{dL_n}{dg} = -\frac{y_n}{g_\theta(x_n)} + \frac{(1-y_n)}{1 - g_\theta(x_n)} \qquad \frac{dL_n}{d\theta} = \frac{dL_n}{dg} \cdot \frac{dg}{d\theta}$$

$$= \frac{(1-y_n)\, g_\theta(x_n) - y_n \cdot (1 - g_\theta(x_n))}{g_\theta(x_n) \cdot (1 - g_\theta(x_n))}$$

$$= \frac{g_\theta(x_n) - y_n}{\cancel{g_\theta(x_n) \cdot (1 - g_\theta(x_n))}} \qquad 1$$

$$g_\theta(x_n) = \sigma(\theta^T x_n) = \frac{1}{1 + \exp(-\theta^T x_n)}$$

$$\frac{dg}{d\theta} = \frac{1}{[1 + \exp(-\theta^T x_n)]^2} (-1) \cdot \exp(-\theta^T x_n) \cdot (-x_n)$$

$$= \frac{\exp(-\theta^T x_n)\, x_n}{[1 + \exp(-\theta^T (x_n))]^2}$$

$$= \quad x_n \quad \cancel{g_\theta(x_n)\cdot(1 - g_\theta(x_n))} \quad 2$$

$$\frac{dL_n}{d\theta} = 1 \times 2 = (g_\theta(x_n) - y_n)\, x_n$$

$$\frac{dL}{d\theta} = \frac{1}{N} \sum_n (g_\theta(x_n) - y_n) \cdot x_n$$