

$$J(\theta) = \left\{ \sum_{i=1}^N y_i \log(\hat{y}_i) + (1-y_i) \log(1-\hat{y}_i) \right\}$$

$$\frac{\partial J(\theta)}{\partial \theta} = -\sum_{i=1}^N \left\{ \frac{y_i}{\hat{y}_i} \frac{\partial \hat{y}_i}{\partial \theta} + \frac{(1-y_i)}{1-\hat{y}_i} \frac{\partial \hat{y}_i}{\partial \theta} \right\} \quad \dots (1)$$

Put (2) in (1)

$$\frac{\partial J(\theta)}{\partial \theta} = -\sum_{i=1}^N \left\{ \frac{y_i}{\hat{y}_i} \left( \frac{\hat{y}_i}{y_i} \right) (1-y_i) * x_i + \frac{(1-y_i)}{1-\hat{y}_i} \left( \frac{\hat{y}_i}{1-y_i} \right) (-1) * (y_i) (1-x_i) \right\} \quad \dots (1)$$

ASIDE

$$\hat{y}_i = \frac{1}{1+e^{-z_i}} = g_\theta(z_i) \quad \text{where} \quad z_i = \alpha^\top \theta$$

$$\begin{aligned} \frac{\partial \hat{y}_i}{\partial \theta} &= g'_\theta(z_i)(1-g_\theta(z_i))x_i + (-g'_\theta(z_i)(1-g_\theta(z_i)))x_i \\ &= -\sum_{i=1}^N \left\{ (y_i - \hat{y}_i) \hat{y}_i x_i \right\} \\ &= -\sum_{i=1}^N \left\{ (y_i - \hat{y}_i) x_i \right\} \end{aligned}$$

$$\begin{aligned} \frac{\partial \hat{y}_i}{\partial \theta} &= g'_\theta(z_i)(1-g_\theta(z_i)) * \frac{\partial z_i}{\partial \theta} \\ &= (y_i)(1-y_i)x_i \quad \dots (2) \end{aligned}$$

(3)