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

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

$\boxed{\text{ASIDE}}$

$$\hat{y}_i = \frac{1}{1+e^{-z}} = \sigma_\theta(z_i) \quad \text{where } z_i = x_i^T \theta$$

$$\frac{\partial \hat{y}_i}{\partial \theta} = \sigma_\theta(z_i)(1-\sigma_\theta(z_i)) * \frac{\partial z_i}{\partial \theta}$$

$$\frac{\partial \hat{y}_i}{\partial \theta} = (\hat{y}_i)(1-\hat{y}_i) * x_i \quad \ldots 2$$

Put 2 in 1

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

$$= -\sum_{i=1}^{N} \left\{ (y_i - y_i \cancel{\hat{y}_i}) x_i + (-\hat{y}_i + y_i \cancel{\hat{y}_i})(x_i) \right.$$

$$= -\sum_{i=1}^{N} \left\{ (y_i - \hat{y}_i) x_i \right\} = \sum_{i=1}^{N} \left\{ \hat{y}_i - y_i \right\} x_i$$

3