

$$\frac{\partial J(\omega)}{\partial \omega_j} = \sum_i \frac{\partial}{\partial \omega_j} \ell(y_i; x_i; \omega)$$

$$\frac{\partial}{\partial \omega_j} \ell(y_i; x_i; \omega) = \frac{\partial}{\partial \omega_j} \left[(1-y_i) \log(1-p(x_i; \omega)) + y_i \log p_1(x_i; \omega) \right]$$

$$= (1-y_i) \frac{1}{1-p_1(x_i; \omega)} \left(- \frac{\partial p_1(x_i; \omega)}{\partial \omega_j} \right) + y_i \frac{1}{p_1(x_i; \omega)} \frac{\partial p_1(x_i; \omega)}{\partial \omega_j}$$

$$= \left[\frac{-(1-y_i)}{1-p_1(x_i; \omega)} + \frac{y_i}{p_1(x_i; \omega)} \right] \frac{\partial p_1(x_i; \omega)}{\partial \omega_j}$$

$$= \left[\frac{y_i - y_i p_1(x_i; \omega) - p_1(x_i; \omega) + y_i p_1(x_i; \omega)}{p_1(x_i; \omega) (1-p_1(x_i; \omega))} \right] \frac{\partial p_1(x_i; \omega)}{\partial \omega_j}$$

$$= \left[\frac{y_i - p_1(x_i; \omega)}{p_1(x_i; \omega) (1-p_1(x_i; \omega))} \right] \frac{\partial p_1(x_i; \omega)}{\partial \omega_j}$$

①

$$P_i(x_i; \omega) = \frac{1}{1 + e^{-\omega \cdot x_i}}$$

$$\frac{\partial P_i(x_i; \omega)}{\partial \omega_j} = \frac{1}{(1 + e^{-\omega \cdot x_i})^2} \frac{\partial}{\partial \omega_j} (1 + e^{-\omega \cdot x_i})$$

$$= \frac{1}{(1 + e^{-\omega \cdot x_i})^2} (e^{-\omega \cdot x_i}) \frac{\partial}{\partial \omega_j} (-\omega \cdot x_j)$$

$$= \frac{e^{-\omega \cdot x_i}}{(1 + e^{-\omega \cdot x_i})^2} \frac{1}{(1 + e^{-\omega \cdot x_i})} x_{ij}$$

$$= P_i(x_i; \omega) (1 - P_i(x_i; \omega)) x_{ij}$$

②

② in ①

$$\frac{\partial}{\partial \omega_j} l(y_i; x_i; \omega) = \left[\frac{(y_i - P_i(x_i; \omega))}{P_i(x_i; \omega) (1 - P_i(x_i; \omega))} \right] P_i(x_i; \omega) (1 - P_i(x_i; \omega)) x_{ij}$$

$$= (y_i - P_i(x_i; \omega)) x_{ij}$$

Over all gradient

$$\frac{\partial J(\omega)}{\partial \omega_j} = \sum_i (y_i - P_i(x_i; \omega)) x_{ij}$$