Maximum Likelihood Estimation.

$$\underset{\theta}{\arg\max} \prod_{1}^{m} P(y^{(i)} \mid x^{(i)}; \theta)$$ → x값이 주어져있을 때, 확률값을 최대화할 수 있는 θ 찾기.

$$P_1^{v_1} \times P_2^{v_2} \times \cdots P_j^{v_j}$$

일어나면 1, 일어나지 않으면 0.

Negative Log-Likelyhood

m개의 data, k개의 class

$$L = \prod_{i=1}^{m} P(y^{(i)} \mid x^{(i)}; \theta) = \prod_{i=1}^{m} \prod_{j=1}^{k} P_j^{(i)V_{ij}},$$

ex)

	$j=1$	$j=2$	$j=3$
$i=1$	1	0	0

$$\Rightarrow \begin{cases} V_{11} = 1 \\ V_{12} = 0 \\ V_{13} = 0 \end{cases}$$

Where $V_{ij} = \begin{cases} 1 & \text{if } y^{(i)} \text{ is label } j \\ 0 & \text{if } y^{(i)} \text{ is not label } j \end{cases}$

Loss를 최소화하기 위해 (-) 붙임.

$$-\log L = -\log \prod_{i=1}^{m} \prod_{j=1}^{k} P_j^{(i)V_{ij}} = \underbrace{-\sum_{i=1}^{m} \sum_{j=1}^{k} V_{ij} \log P_j^{(i)}}_{= l}$$

최소화해야 할 값.
최소화하는 W_j 값을 찾아야 함.

$\frac{\partial l}{\partial P_j}$ 해야하는데, 여기서 $P_j^{(i)} = \frac{e^{z_j^{(i)}}}{\sum_{j=1}^{k} e^{z_j^{(i)}}}$

$$\frac{\partial l}{\partial z_i} = -\sum_{j=1}^{k} V_{ij} \frac{\partial \log P_j}{\partial z_i} = -\sum_{j=1}^{k} V_{ij} \frac{1}{P_j} \cdot \frac{\partial P_j}{\partial z_i} \quad \left(\because \frac{d \log_e f(x)}{dx} = \frac{1}{f(x)} \cdot \frac{df(x)}{dx}\right)$$

$$\frac{\partial P_j}{\partial z_c} = \frac{\partial \frac{e^{z_j}}{\sum_{k=1}^{k} e^{z_k}}}{\partial z_c}$$ (여기서, $g_j = e^{z_j}$, $h_j = \sum_{k=1}^{k} e^{z_k}$ 로 둠)

i) if $c = j$ (여기서 c와 j는 임의의 값)

$$\frac{\partial P_j}{\partial z_c} = \frac{e^{z_j} \cdot h_j - e^{z_c} e^{z_j}}{[h_j]^2} = \frac{e^{z_j}}{h_j} \cdot \frac{h_j - e^{z_c}}{h_j} = P_j (1 - P_j) \quad \left(\because P_j = \frac{e^{z_j}}{\sum_{j=1}^{k} e^{z_j}}\right)$$

ii) if $c \neq j$

$$\frac{\partial p_j}{\partial z_c} = \frac{\partial \dfrac{e^{z_j}}{\sum_{k=1}^{K} e^{z_k}}}{\partial z_c} = \frac{0 - e^{z_c} e^{z_j}}{[h_j]^2} = -\frac{e^{z_j}}{h_j} \cdot \frac{e^{z_c}}{h_j} = -p_j p_c$$

$$\frac{\partial l}{\partial z_i} = -\sum_{j=1}^{K} y_{ij} \frac{1}{p_j} \cdot \frac{\partial p_j}{\partial z_i} = \underbrace{-\frac{y_i}{p_i} \frac{\partial p_i}{\partial z_i}}_{j=i} \underbrace{- \sum_{j\neq i}^{K} \frac{y_j}{p_j} \cdot \frac{\partial p_j}{\partial z_i}}_{j \neq i \text{인 모든 경우}}$$

$$= -\frac{y_i}{p_i} \cdot p_i(1-p_i) - \sum_{j\neq i}^{K} \frac{y_j}{p_j}(-p_j \cdot p_i)$$

$$= -y_i + y_i p_i + \sum_{j\neq i}^{K} y_j p_i = -y_i + \sum_{j=1}^{K} y_j p_i = p_i - y_i$$

$\sum_{j\neq i}^{K} y_j p_i$ 에서 $j=i$인 경우

Update weights.

$$W_{kj} = W_{kj} - \alpha \frac{\partial l}{\partial W_{kj}} = W_{kj} - \alpha \cdot \frac{\partial l}{\partial z_k} \cdot \frac{\partial z_k}{\partial W_{kj}}$$

$$= W_{kj} - \alpha (p_k - y_k) x_j = W_{kj} + \alpha (y_k - p_k) x_j$$

$$\Rightarrow W_{kj} = W_{kj} + \alpha \sum_{i=1}^{n} \left(y_k^{(i)} - p_k^{(i)}\right) x_j^{(i)}$$

$$W = \begin{bmatrix} W_{10} & W_{11} & \cdots & W_{1j} \\ W_{20} & W_{21} & \cdots & W_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ W_{k0} & W_{k1} & \cdots & W_{kj} \end{bmatrix}$$

where $\begin{cases} k & \text{index of class} \\ j & \text{index of feature.} \end{cases}$