Lecture.7
Linear/Logistic Regression(2)

- Linear Regression

Model and Tensors

$$J \uparrow$$

$$\boxed{J = \mathscr{L}(\vec{y}, \vec{\hat{y}})}$$

$$\uparrow \vec{\hat{y}}$$

$$\boxed{\vec{\hat{y}} = X^T \cdot \vec{w} + b}$$

$$\uparrow X \qquad \vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix}, \; b$$

$$J = \frac{1}{N}\sum_{i=1}^{N} J_0^{(i)}$$

$$\vec{J_0} = \begin{pmatrix} \left(y^{(1)} - \hat{y}^{(1)}\right)^2 \\ \left(y^{(2)} - \hat{y}^{(2)}\right)^2 \\ \vdots \\ \left(y^{(N)} - \hat{y}^{(N)}\right)^2 \end{pmatrix} = \begin{pmatrix} J_0^{(1)} \\ J_0^{(2)} \\ \vdots \\ J_0^{(N)} \end{pmatrix}$$

$$\vec{\hat{y}} = X^T \cdot \vec{w} + b = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix} \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_{l_I} \end{pmatrix} + b = \begin{pmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(N)} \end{pmatrix}$$

$$X^T = \begin{pmatrix} \longleftarrow & \left(\vec{x}^{(1)}\right)^T & \longrightarrow \\ \longleftarrow & \left(\vec{x}^{(2)}\right)^T & \longrightarrow \\ & \vdots & \\ \longleftarrow & \left(\vec{x}^{(N)}\right)^T & \longrightarrow \end{pmatrix} = \begin{pmatrix} x_1^{(1)} & x_2^{(1)} & \dots & x_{l_I}^{(1)} \\ x_1^{(2)} & x_2^{(2)} & \dots & x_{l_I}^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ x_1^{(N)} & x_2^{(N)} & \dots & x_{l_I}^{(N)} \end{pmatrix}$$

$$\therefore \frac{\partial J}{\partial \vec{w}} = \frac{\partial J}{\partial \vec{J_0}} \cdot \frac{\partial \vec{J_0}}{\partial \vec{\hat{y}}} \cdot \frac{\partial \vec{\hat{y}}}{\partial \vec{w}} = (1, l_I)$$

$$\cdot \frac{\partial J}{\partial b} = \frac{\partial J}{\partial \vec{J_0}} \cdot \frac{\partial \vec{J_0}}{\partial \vec{\hat{y}}} \cdot \frac{\partial \vec{\hat{y}}}{\partial b} = (1,1)$$

$$J = \frac{1}{N}\sum_{i=1}^{N} J_0^{(i)} \rightsquigarrow J = \frac{1}{N}\sum_{i=1}^{N} J_0^{(i)} \qquad \frac{\partial J}{\partial \vec{J_0}} = \left(\frac{1}{N}, \dots, \frac{1}{N}\right) \in \mathbb{R}^{1\times N}$$

↳ Multivariate

$$J_0 = \mathcal{L}(\vec{y}, \vec{\hat{y}}) \rightsquigarrow \vec{J_0} = \vec{\mathcal{L}}(\vec{y}, \vec{\hat{y}}) = \begin{pmatrix} J_0^{(1)} \\ J_0^{(2)} \\ \vdots \\ J_0^{(N)} \end{pmatrix} \qquad \frac{\partial \vec{J_0}}{\partial \vec{\hat{y}}} = -2\begin{pmatrix} y^{(1)}-\hat{y}^{(1)} & 0 & \dots & 0 \\ 0 & y^{(2)}-\hat{y}^{(2)} & & 0 \\ \vdots & & \ddots & \\ 0 & 0 & \dots & y^{(N)}-\hat{y}^{(N)} \end{pmatrix}$$

↳ Binary Element-wise

$$\vec{\hat{y}} = X^T \cdot \vec{w} + b \qquad \vec{\hat{y}} = X^T \cdot \vec{w} + b = \begin{pmatrix} \longleftarrow (\vec{x}^{(1)})^T \longrightarrow \\ \vdots \\ \longleftarrow (\vec{x}^{(N)})^T \longrightarrow \end{pmatrix}\begin{pmatrix} \uparrow \\ \vec{w} \\ \downarrow \end{pmatrix} + b$$

↳ Vector function

$$\uparrow X$$

$$\boxed{D.S}$$

$$\begin{pmatrix} \uparrow & \uparrow & & \uparrow \\ \vec{x}^{(1)} & \vec{x}^{(2)} & \dots & \vec{x}^{(N)} \\ \downarrow & \downarrow & & \downarrow \end{pmatrix}$$

$$\cdot \frac{\partial \vec{\hat{y}}}{\partial \vec{w}} = \begin{pmatrix} \longleftarrow (\vec{x}^{(1)})^T \longrightarrow \\ \vdots \\ \longleftarrow (\vec{x}^{(N)})^T \longrightarrow \end{pmatrix} = X^T \qquad \cdot \frac{\partial \vec{\hat{y}}}{\partial b} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

$$\frac{\partial \hat{y}^{(i)}}{\partial \vec{w}} \quad \hat{y}^{(i)} = (\vec{x}^{(i)})^T \cdot \vec{w} + b \qquad (\vec{x}^{(i)})^T$$

$$\frac{\partial J}{\partial \vec{J_0}}, \; \frac{\partial \vec{J_0}}{\partial \vec{\hat{y}}}, \; \frac{\partial \vec{\hat{y}}}{\partial \vec{w}}, \; \frac{\partial \vec{\hat{y}}}{\partial b}$$

$$\frac{\partial J}{\partial \vec{\hat{y}}} = \frac{\partial J}{\partial \vec{J_0}} \cdot \frac{\partial \vec{J_0}}{\partial \vec{\hat{y}}} = -\frac{2}{N}\left(y^{(1)}-\hat{y}^{(1)} \; \dots \; y^{(N)}-\hat{y}^{(N)}\right)$$

$$\frac{\partial J}{\partial \vec{w}} = \frac{\partial J}{\partial \vec{\hat{y}}} \cdot \frac{\partial \vec{\hat{y}}}{\partial \vec{w}} = -\frac{2}{N}\left(y^{(1)}-\hat{y}^{(1)} \; \dots \; y^{(N)}-\hat{y}^{(N)}\right)\begin{pmatrix} \longleftarrow (\vec{x}^{(1)})^T \longrightarrow \\ \vdots \\ \longleftarrow (\vec{x}^{(N)})^T \longrightarrow \end{pmatrix} = -\frac{2}{N}\left(\sum_{i=1}^{N} x_1^{(i)}(y^{(i)}-\hat{y}^{(i)}) \; \sum_{i=1}^{N} x_2^{(i)}(y^{(i)}-\hat{y}^{(i)}) \; \dots \; \sum_{i=1}^{N} x_{l_I}^{(i)}(y^{(i)}-\hat{y}^{(i)})\right)$$

$$\frac{\partial J}{\partial w_k} = -\frac{2}{N}\sum_{i=1}^{N} x_k^{(i)}(y^{(i)}-\hat{y}^{(i)})$$

- 각 data의 sample의 Gradient를 가져오는 데 필요
- 각 data에 sample의 Gradient를 update시키는 방법, 각 번을 종합해서 하는 방식 (Minibatch로 사용할 때)

$$J = (y-\hat{y})^2 = (y - (\vec{x}^T\vec{w}+b))^2$$

$$\frac{\partial J}{\partial \vec{w}} = -2 \cdot \vec{x}^T \cdot (y-\hat{y})$$

$$\frac{\partial J}{\partial b} = \frac{\partial J}{\partial \vec{\hat{y}}} \cdot \frac{\partial \vec{\hat{y}}}{\partial b} = -\frac{2}{N}\left(y^{(1)}-\hat{y}^{(1)} \; \dots \; y^{(N)}-\hat{y}^{(N)}\right)\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = -\frac{2}{N}\sum_{i=1}^{N}(y^{(i)}-\hat{y}^{(i)})$$