Matrix Differentiation [Detailed Derivation]

Variables

- A - Input Matrix

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} \\ a_{21} & a_{22} & \dots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} \end{bmatrix}$$

- W - Weight Matrix

$$W = \begin{bmatrix} w_{11} \\ w_{21} \\ \vdots \\ w_{j1} \end{bmatrix}$$

- $\hat{Y}$ - Y predict

$$\hat{Y} = A.W$$

$$\hat{Y} = \begin{bmatrix} \hat{y}_{11} \\ \hat{y}_{21} \\ \vdots \\ \hat{y}_{i1} \end{bmatrix}$$

$$\hat{Y} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} \\ a_{21} & a_{22} & \dots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} \end{bmatrix} \cdot \begin{bmatrix} w_{11} \\ w_{21} \\ \vdots \\ w_{j1} \end{bmatrix}$$

- Y - Y true

$$Y = \begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{i1} \end{bmatrix}$$

Expanding $\hat{Y}$

$$\hat{Y} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1j} \\ a_{21} & a_{22} & \dots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \dots & a_{ij} \end{bmatrix} \cdot \begin{bmatrix} w_{11} \\ w_{21} \\ \vdots \\ w_{j1} \end{bmatrix}$$

$$= \begin{bmatrix} a_{11}w_{11} + a_{12}w_{21} + \dots + a_{1j}w_{j1} \\ a_{21}w_{11} + a_{22}w_{21} + \dots + a_{2j}w_{j1} \\ \vdots \\ a_{i1}w_{11} + a_{i2}w_{21} + \dots + a_{ij}w_{j1} \end{bmatrix}$$

---

loss

$$loss = \frac{1}{N} \sum (Y - \hat{Y})^2$$

$$= \frac{1}{N} \begin{pmatrix} (y_{11} - a_{11}w_{11} - a_{12}w_{21} - \dots - a_{1j}w_{j1})^2 \\ +(y_{21} - a_{21}w_{11} - a_{22}w_{21} - \dots - a_{2j}w_{j1})^2 \\ \vdots \\ +(y_{i1} - a_{i1}w_{11} - a_{i2}w_{21} - \dots - a_{ij}w_{j1})^2 \end{pmatrix}$$

> Here N is number of rows in input matrix A

Finding δW

$$\frac{d(loss)}{d(w_{j1})} = \frac{1}{N} \frac{d}{d(w_{j1})} \begin{pmatrix} (y_{11} - a_{11}w_{11} - a_{12}w_{21} - \dots - a_{1j}w_{j1})^2 \\ +(y_{21} - a_{21}w_{11} - a_{22}w_{21} - \dots - a_{2j}w_{j1})^2 \\ \vdots \\ +(y_{i1} - a_{i1}w_{11} - a_{i2}w_{21} - \dots - a_{ij}w_{j1})^2 \end{pmatrix}$$

$$\frac{d(loss)}{d(w_{j1})} = \frac{1}{N} \begin{pmatrix} \frac{d}{d(w_{j1})}(y_{11} - a_{11}w_{11} - a_{12}w_{21} - \dots - a_{1j}w_{j1})^2 \\ +\frac{d}{d(w_{j1})}(y_{21} - a_{21}w_{11} - a_{22}w_{21} - \dots - a_{2j}w_{j1})^2 \\ \vdots \\ +\frac{d}{d(w_{j1})}(y_{i1} - a_{i1}w_{11} - a_{i2}w_{21} - \dots - a_{ij}w_{j1})^2 \end{pmatrix}$$

$$\frac{d(loss)}{d(w_{j1})} = \frac{1}{N} \begin{pmatrix} 2(y_{11} - a_{11}w_{11} - a_{12}w_{21} - \dots - a_{1j}w_{j1})(-a_{1j}) \\ +2(y_{21} - a_{21}w_{11} - a_{22}w_{21} - \dots - a_{2j}w_{j1})(-a_{2j}) \\ \vdots \\ +2(y_{i1} - a_{i1}w_{11} - a_{i2}w_{21} - \dots - a_{ij}w_{j1})(-a_{ij}) \end{pmatrix}$$

$$\frac{d(loss)}{d(w_{j1})} = -\frac{2}{N}\left(\begin{matrix} (y_{11} - a_{11}w_{11} - a_{12}w_{21} - \cdots - a_{1j}w_{j1})(a_{1j}) \\ +(y_{21} - a_{21}w_{11} - a_{22}w_{21} - \cdots - a_{2j}w_{j1})(a_{2j}) \\ \vdots \\ +(y_{i1} - a_{i1}w_{11} - a_{i2}w_{21} - \cdots - a_{ij}w_{j1})(a_{ij}) \end{matrix}\right)$$

$$\frac{d(loss)}{d(w_{j1})} = -\frac{2}{N}\begin{bmatrix} a_{1j} & a_{2j} & \ldots & a_{ij} \end{bmatrix} \cdot \begin{bmatrix} (y_{11} - a_{11}w_{11} - a_{12}w_{21} - \cdots - a_{1j}w_{j1}) \\ (y_{21} - a_{21}w_{11} - a_{22}w_{21} - \cdots - a_{2j}w_{j1}) \\ \vdots \\ (y_{i1} - a_{i1}w_{11} - a_{i2}w_{21} - \cdots - a_{ij}w_{j1}) \end{bmatrix}$$

$$\frac{d(loss)}{d(w_{j1})} = -\frac{2}{N}\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{ij} \end{bmatrix}^T \cdot \begin{bmatrix} (y_{11} - a_{11}w_{11} - a_{12}w_{21} - \cdots - a_{1j}w_{j1}) \\ (y_{21} - a_{21}w_{11} - a_{22}w_{21} - \cdots - a_{2j}w_{j1}) \\ \vdots \\ (y_{i1} - a_{i1}w_{11} - a_{i2}w_{21} - \cdots - a_{ij}w_{j1}) \end{bmatrix}$$

$$\frac{d(loss)}{d(w_{j1})} = -\frac{2}{N}\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{ij} \end{bmatrix}^T \cdot \left[\begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{i1} \end{bmatrix} - \begin{bmatrix} a_{11}w_{11} + a_{12}w_{21} + \cdots + a_{1j}w_{j1} \\ a_{21}w_{11} + a_{22}w_{21} + \cdots + a_{2j}w_{j1} \\ \vdots \\ a_{i1}w_{11} + a_{i2}w_{21} + \cdots + a_{ij}w_{j1} \end{bmatrix}\right]$$

$$\frac{d(loss)}{d(w_{j1})} = -\frac{2}{N}\begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{ij} \end{bmatrix}^T \cdot \left[\begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{i1} \end{bmatrix} - \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1j} \\ a_{21} & a_{22} & \ldots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{ij} \end{bmatrix} \cdot \begin{bmatrix} w_{11} \\ w_{21} \\ \vdots \\ w_{j1} \end{bmatrix}\right]$$

$$\delta W = -\frac{2}{N}\begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1j} \\ a_{21} & a_{22} & \ldots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{ij} \end{bmatrix}^T \cdot \left[\begin{bmatrix} y_{11} \\ y_{21} \\ \vdots \\ y_{i1} \end{bmatrix} - \begin{bmatrix} a_{11} & a_{12} & \ldots & a_{1j} \\ a_{21} & a_{22} & \ldots & a_{2j} \\ \vdots & \vdots & \ddots & \vdots \\ a_{i1} & a_{i2} & \ldots & a_{ij} \end{bmatrix} \cdot \begin{bmatrix} w_{11} \\ w_{21} \\ \vdots \\ w_{j1} \end{bmatrix}\right]$$

grad: $\delta W = -\frac{2}{N}A^T.(Y - A.W)$
