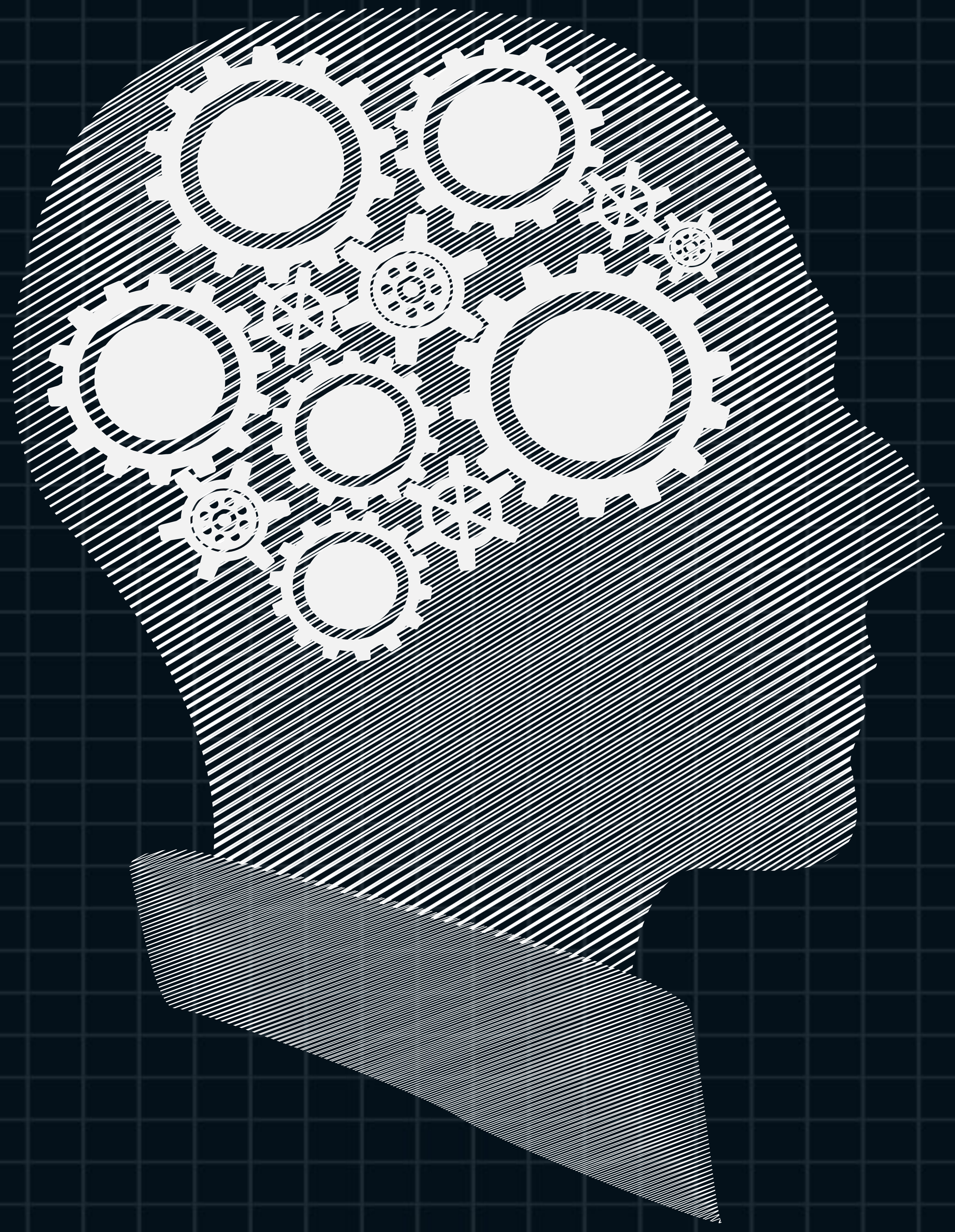


Backpropagation and Jacobian Matrices

Lecture.10
Expanded Jacobians in Deep Learning

Mean Squared Error

$$J = \frac{1}{N}\sum_{i=1}^{N}\left(y^{(i)} - \hat{y}^{(i)}\right)^2$$

$$\hat{Y} = \begin{pmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(N)} \end{pmatrix} \in \mathbb{R}^{N\times 1}, \quad Y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{pmatrix} \in \mathbb{R}^{N\times 1}$$

$$\frac{\partial J}{\partial \hat{y}^{(i)}} = -\frac{2}{N}\left(y^{(i)} - \hat{y}^{(i)}\right)$$

$$d\hat{Y} = \begin{pmatrix} d\hat{y}^{(1)} \\ d\hat{y}^{(2)} \\ \vdots \\ d\hat{y}^{(N)} \end{pmatrix} = \begin{pmatrix} -\frac{2}{N}\left(y^{(1)} - \hat{y}^{(1)}\right) \\ -\frac{2}{N}\left(y^{(2)} - \hat{y}^{(2)}\right) \\ \vdots \\ -\frac{2}{N}\left(y^{(N)} - \hat{y}^{(N)}\right) \end{pmatrix} \in \mathbb{R}^{N\times 1}$$

$$d\hat{Y} = -2/N * (Y - \hat{Y})$$

Binary Cross Entropy Error

$$J = -\frac{1}{N}\sum_{i=1}^{N}\left[y^{(i)}log\left(\hat{y}^{(i)}\right) + \left(1 - y^{(i)}\right)log\left(1 - \hat{y}^{(i)}\right)\right]$$

$$\hat{Y} = \begin{pmatrix} \hat{y}^{(1)} \\ \hat{y}^{(2)} \\ \vdots \\ \hat{y}^{(N)} \end{pmatrix} \in \mathbb{R}^{N\times 1},\ \ Y = \begin{pmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(N)} \end{pmatrix} \in \mathbb{R}^{N\times 1}$$

$$d\hat{y}^{(i)} = \frac{1}{N}\frac{\hat{y}^{(i)} - y^{(i)}}{\hat{y}^{(i)}\left(1 - \hat{y}^{(i)}\right)}$$

$$d\hat{Y} = \begin{pmatrix} d\hat{y}^{(1)} \\ d\hat{y}^{(2)} \\ \vdots \\ d\hat{y}^{(N)} \end{pmatrix} = \frac{1}{N}\begin{pmatrix} \frac{\hat{y}^{(1)} - y^{(1)}}{\hat{y}^{(1)}\left(1 - \hat{y}^{(1)}\right)} \\ \frac{\hat{y}^{(2)} - y^{(2)}}{\hat{y}^{(2)}\left(1 - \hat{y}^{(2)}\right)} \\ \vdots \\ \frac{\hat{y}^{(N)} - y^{(N)}}{\hat{y}^{(N)}\left(1 - \hat{y}^{(N)}\right)} \end{pmatrix} \in \mathbb{R}^{N\times 1}$$

$$d\hat{Y} = \frac{1}{N} * \left(\hat{Y} - Y\right) / \left(\hat{Y}\left(1 - \hat{Y}\right)\right)$$

Categorical Cross Entropy Error

$$J = -\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_k^{(i)} log\left(\hat{y}^{(i)}\right)$$

$$\hat{Y} = \begin{pmatrix} \hat{y}_1^{(1)} & \hat{y}_2^{(1)} & \dots & \hat{y}_K^{(1)} \\ \hat{y}_1^{(2)} & \hat{y}_2^{(2)} & \dots & \hat{y}_K^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ \hat{y}_1^{(N)} & \hat{y}_2^{(N)} & \dots & \hat{y}_K^{(N)} \end{pmatrix} \in \mathbb{R}^{N\times K},\ Y = \begin{pmatrix} y_1^{(1)} & y_2^{(1)} & \dots & y_K^{(1)} \\ y_1^{(2)} & y_2^{(2)} & \dots & y_K^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(N)} & y_2^{(N)} & \dots & y_K^{(N)} \end{pmatrix} \in \mathbb{R}^{N\times K}$$

$$\begin{aligned}
\frac{\partial J}{\partial \hat{y}_\beta^{(\alpha)}} &= \frac{\partial}{\partial \hat{y}_\beta^{(\alpha)}}\left[-\frac{1}{N}\sum_{i=1}^{N}\sum_{k=1}^{K} y_k^{(i)} log\left(\hat{y}_k^{(i)}\right)\right] \\
&= -\frac{1}{N}\frac{\partial}{\partial \hat{y}_\beta^{(\alpha)}}\left[\sum_{i=1}^{N}\sum_{k=1}^{K} y_k^{(i)} log\left(\hat{y}_k^{(i)}\right)\right] \\
&= -\frac{1}{N}\frac{\partial}{\partial \hat{y}_\beta^{(\alpha)}}\left[\sum_{k=1}^{K} y_k^{(\alpha)} log\left(\hat{y}_k^{(\alpha)}\right)\right] \\
&= -\frac{1}{N}\frac{\partial}{\partial \hat{y}_\beta^{(\alpha)}}\left[y_\beta^{(\alpha)} log\left(\hat{y}_\beta^{(\alpha)}\right)\right] \\
&= -\frac{1}{N}\cdot\frac{y_\beta^{(\alpha)}}{\hat{y}_\beta^{(\alpha)}}
\end{aligned}$$

$$\frac{\partial J}{\partial \hat{y}_j^{(i)}} = -\frac{1}{N}\frac{y_j^{(i)}}{\hat{y}_j^{(i)}}$$

$$d\hat{Y} = \begin{pmatrix} d\,\vec{\hat{y}}^{(1)} \\ d\,\vec{\hat{y}}^{(2)} \\ \vdots \\ d\,\vec{\hat{y}}^{(N)} \end{pmatrix} = \begin{pmatrix} d\hat{y}_1^{(1)} & d\hat{y}_2^{(1)} & \dots & d\hat{y}_K^{(1)} \\ d\hat{y}_1^{(2)} & d\hat{y}_2^{(2)} & \dots & d\hat{y}_K^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ d\hat{y}_1^{(N)} & d\hat{y}_2^{(N)} & \dots & d\hat{y}_K^{(N)} \end{pmatrix} = -\frac{1}{N}\begin{pmatrix} \frac{y_1^{(1)}}{\hat{y}_1^{(1)}} & \frac{y_2^{(1)}}{\hat{y}_2^{(1)}} & \dots & \frac{y_K^{(1)}}{\hat{y}_K^{(1)}} \\ \frac{y_1^{(2)}}{\hat{y}_1^{(2)}} & \frac{y_2^{(2)}}{\hat{y}_2^{(2)}} & \dots & \frac{y_K^{(2)}}{\hat{y}_K^{(2)}} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{y_1^{(N)}}{\hat{y}_1^{(N)}} & \frac{y_2^{(N)}}{\hat{y}_2^{(N)}} & \dots & \frac{y_K^{(N)}}{\hat{y}_K^{(N)}} \end{pmatrix}$$

$$d\hat{Y} = -\frac{1}{N} * Y/\hat{Y}$$

Total Derivative within Softmax

$$\hat{y}_j^{(i)} = \frac{e^{l_j^{(i)}}}{\sum_{k=1}^{K} e^{l_k^{(i)}}}$$

$$L = \begin{pmatrix} \boxed{l_1^{(1)}} & l_2^{(1)} & \cdots & l_K^{(1)} \\ l_1^{(2)} & \boxed{l_2^{(2)}} & \cdots & l_K^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ l_1^{(N)} & l_2^{(N)} & \cdots & \boxed{l_K^{(N)}} \end{pmatrix} \qquad \hat{Y} = \begin{pmatrix} \boxed{\hat{y}_1^{(1)}} & \boxed{\hat{y}_2^{(1)}} & \cdots & \boxed{\hat{y}_K^{(1)}} \\ \boxed{\hat{y}_1^{(2)}} & \boxed{\hat{y}_2^{(2)}} & \cdots & \boxed{\hat{y}_K^{(2)}} \\ \vdots & \vdots & \ddots & \vdots \\ \boxed{\hat{y}_1^{(N)}} & \boxed{\hat{y}_2^{(N)}} & \cdots & \boxed{\hat{y}_K^{(N)}} \end{pmatrix}$$

$$\hat{y}_j^{(1)} = \frac{e^{l_j^{(1)}}}{\sum_{k=1}^{K} e^{l_k^{(1)}}}$$

$$\hat{y}_j^{(2)} = \frac{e^{l_j^{(2)}}}{\sum_{k=1}^{K} e^{l_k^{(2)}}}$$

$$\hat{y}_j^{(N)} = \frac{e^{l_j^{(N)}}}{\sum_{k=1}^{K} e^{l_k^{(N)}}}$$

$$\frac{\partial J}{\partial l_j^{(i)}} = \frac{\partial J}{\partial \hat{y}_1^{(i)}} \frac{\partial \hat{y}_1^{(i)}}{\partial l_j^{(i)}} + \frac{\partial J}{\partial \hat{y}_2^{(i)}} \frac{\partial \hat{y}_2^{(i)}}{\partial l_j^{(i)}} + \dots + \frac{\partial J}{\partial \hat{y}_j^{(i)}} \frac{\partial \hat{y}_j^{(i)}}{\partial l_j^{(i)}} + \dots + \frac{\partial J}{\partial \hat{y}_K^{(i)}} \frac{\partial \hat{y}_K^{(i)}}{\partial l_j^{(i)}}$$

$$= \sum_{k=1}^{K} \frac{\partial J}{\partial \hat{y}_k^{(i)}} \frac{\partial \hat{y}_k^{(i)}}{\partial l_j^{(i)}}$$

- Softmax and Expanded Jacobians

Total Derivative Calculation

$$\frac{\partial J}{\partial l_j^{(i)}} = \frac{\partial J}{\partial \hat{y}_1^{(i)}}\frac{\partial \hat{y}_1^{(i)}}{\partial l_j^{(i)}} + \frac{\partial J}{\partial \hat{y}_2^{(i)}}\frac{\partial \hat{y}_2^{(i)}}{\partial l_j^{(i)}} + \dots + \frac{\partial J}{\partial \hat{y}_j^{(i)}}\frac{\partial \hat{y}_j^{(i)}}{\partial l_j^{(i)}} + \dots + \frac{\partial J}{\partial \hat{y}_K^{(i)}}\frac{\partial \hat{y}_K^{(i)}}{\partial l_j^{(i)}}$$

$$= d\hat{y}_1^{(i)}\frac{\partial \hat{y}_1^{(i)}}{\partial l_j^{(i)}} + d\hat{y}_2^{(i)}\frac{\partial \hat{y}_2^{(i)}}{\partial l_j^{(i)}} + \dots + d\hat{y}_j^{(i)}\frac{\partial \hat{y}_j^{(i)}}{\partial l_j^{(i)}} + \dots + d\hat{y}_K^{(i)}\frac{\partial \hat{y}_K^{(i)}}{\partial l_j^{(i)}}$$

$$= d\hat{y}_1^{(i)} \cdot \left(-\hat{y}_1^{(i)}\hat{y}_j^{(i)}\right) + d\hat{y}_2^{(i)} \cdot \left(-\hat{y}_2^{(i)}\hat{y}_j^{(i)}\right) + \dots + d\hat{y}_j^{(i)} \cdot \left(\hat{y}_j^{(i)}\left(1-\hat{y}_j^{(i)}\right)\right) + \dots + d\hat{y}_K^{(i)} \cdot \left(-\hat{y}_K^{(i)}\hat{y}_j^{(i)}\right)$$

$$= d\hat{y}_j^{(i)}\hat{y}_j^{(i)} + d\hat{y}_1^{(i)} \cdot \left(-\hat{y}_1^{(i)}\hat{y}_j^{(i)}\right) + d\hat{y}_2^{(i)} \cdot \left(-\hat{y}_2^{(i)}\hat{y}_j^{(i)}\right) + \dots + d\hat{y}_j^{(i)} \cdot \left(-\hat{y}_j^{(i)}\hat{y}_j^{(i)}\right) + \dots + d\hat{y}_K^{(i)} \cdot \left(-\hat{y}_K^{(i)}\hat{y}_j^{(i)}\right)$$

$$= d\hat{y}_j^{(i)}\hat{y}_j^{(i)} - \hat{y}_j^{(i)}\sum_{k=1}^{K} d\hat{y}_k^{(i)} \cdot \hat{y}_k^{(i)}$$

$$= -\frac{1}{N}\left[\frac{y_j^{(i)}}{\hat{y}_j^{(i)}} \cdot \hat{y}_j^{(i)} - \hat{y}_j^{(i)}\sum_{k=1}^{K}\frac{y_k^{(i)}}{\hat{y}_k^{(i)}} \cdot \hat{y}_k^{(i)}\right]$$

$$= -\frac{1}{N}\left[y_j^{(i)} - \hat{y}_j^{(i)}\sum_{k=1}^{K} y_k^{(i)}\right]$$

$$= -\frac{1}{N}\left[y_j^{(i)} - \hat{y}_j^{(i)}\right]$$

$$\frac{\partial J}{\partial \hat{y}_j^{(i)}} = -\frac{1}{N}\frac{y_j^{(i)}}{\hat{y}_j^{(i)}} \qquad \frac{\partial \hat{y}_\alpha^{(i)}}{\partial l_\beta^{(i)}} = \begin{cases} \hat{y}_\alpha^{(i)}(1-\hat{y}_\alpha^{(i)}), & if\ \alpha = \beta \\ -\hat{y}_\alpha^{(i)}\hat{y}_\beta^{(i)}, & if\ \alpha \neq \beta \end{cases}$$

Expanded Jacobians at Softmax

$$dl_j^{(i)} = -\frac{1}{N}\left[y_j^{(i)} - \hat{y}_j^{(i)}\right]$$

$$dL = \begin{pmatrix} dl_1^{(1)} & dl_2^{(1)} & \dots & dl_K^{(1)} \\ dl_1^{(2)} & dl_2^{(2)} & \dots & dl_K^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ dl_1^{(N)} & dl_2^{(N)} & \dots & dl_K^{(N)} \end{pmatrix} = -\frac{1}{N}\begin{pmatrix} y_1^{(1)} - \hat{y}_1^{(1)} & y_2^{(1)} - \hat{y}_2^{(1)} & \dots & y_K^{(1)} - \hat{y}_K^{(1)} \\ y_1^{(2)} - \hat{y}_1^{(2)} & y_2^{(2)} - \hat{y}_2^{(2)} & \dots & y_K^{(2)} - \hat{y}_K^{(2)} \\ \vdots & \vdots & \ddots & \vdots \\ y_1^{(N)} - \hat{y}_1^{(N)} & y_2^{(N)} - \hat{y}_2^{(N)} & \dots & y_K^{(N)} - \hat{y}_K^{(N)} \end{pmatrix}$$

$$dL = -\frac{1}{N} * (Y - \hat{Y})$$

- Affine Function and Expanded Jacobians

Matrix Multiplication Review

$$A \in \mathbb{R}^{\alpha \times \beta},\ B \in \mathbb{R}^{\beta \times \gamma},\ C \in \mathbb{R}^{\alpha \times \gamma}$$

$$C = AB \qquad c_{ij} = \sum_{k=1}^{\beta} a_{ik} b_{kj}$$

$$Row_i(A) \sim Row_i(C)$$

$$Col_j(B) \sim Col_j(C)$$

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1\beta} \\ a_{21} & a_{22} & \dots & a_{2\beta} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{i\beta} \\ \vdots & \vdots & & \vdots \\ a_{\alpha 1} & a_{\alpha 2} & \dots & a_{\alpha\beta} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1j} & \dots & b_{1\gamma} \\ b_{21} & b_{22} & \dots & b_{2j} & \dots & b_{2\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{\beta 1} & b_{\beta 2} & \dots & b_{\beta j} & \dots & b_{\beta\gamma} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1j} & \dots & c_{1\gamma} \\ c_{21} & c_{22} & \dots & c_{2j} & \dots & c_{2\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{i1} & c_{i2} & \dots & c_{ij} & \dots & c_{i\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{\alpha 1} & c_{\alpha 2} & \dots & c_{\alpha j} & \dots & c_{\alpha\gamma} \end{pmatrix}$$

- Affine Function and Expanded Jacobians

Matrix Multiplication Review

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1\beta} \\ a_{21} & a_{22} & \dots & a_{2\beta} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{i\beta} \\ \vdots & \vdots & & \vdots \\ a_{\alpha 1} & a_{\alpha 2} & \dots & a_{\alpha\beta} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1j} & \dots & b_{1\gamma} \\ b_{21} & b_{22} & \dots & b_{2j} & \dots & b_{2\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{\beta 1} & b_{\beta 2} & \dots & b_{\beta j} & \dots & b_{\beta\gamma} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1j} & \dots & c_{1\gamma} \\ c_{21} & c_{22} & \dots & c_{2j} & \dots & c_{2\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{i1} & c_{i2} & \dots & c_{ij} & \dots & c_{i\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{\alpha 1} & c_{\alpha 2} & \dots & c_{\alpha j} & \dots & c_{\alpha\gamma} \end{pmatrix}$$

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1\beta} \\ a_{21} & a_{22} & \dots & a_{2\beta} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{i\beta} \\ \vdots & \vdots & & \vdots \\ a_{\alpha 1} & a_{\alpha 2} & \dots & a_{\alpha\beta} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1j} & \dots & b_{1\gamma} \\ b_{21} & b_{22} & \dots & b_{2j} & \dots & b_{2\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{\beta 1} & b_{\beta 2} & \dots & b_{\beta j} & \dots & b_{\beta\gamma} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1j} & \dots & c_{1\gamma} \\ c_{21} & c_{22} & \dots & c_{2j} & \dots & c_{2\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{i1} & c_{i2} & \dots & c_{ij} & \dots & c_{i\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{\alpha 1} & c_{\alpha 2} & \dots & c_{\alpha j} & \dots & c_{\alpha\gamma} \end{pmatrix}$$

- Affine Function and Expanded Jacobians

Matrix Multiplication Review

$$M = (m_{ij}) \quad M^T = \left((m_T)_{ij}\right)$$

$$m_{ij} = (m_T)_{ji}$$

$$Row_i(M) = (m_{i1} \quad m_{i2} \quad \dots \quad m_{i\beta})$$

$$Col_j(M) = \begin{pmatrix} m_{1j} \\ m_{2j} \\ \vdots \\ m_{\alpha j} \end{pmatrix}$$

- Affine Function and Expanded Jacobians

Related Tensors(Matrix Multiplication)

$$C = A \cdot B$$

$$A \in \mathbb{R}^{\alpha\times\beta},\ B \in \mathbb{R}^{\beta\times\gamma},\ C \in \mathbb{R}^{\alpha\times\gamma}$$

$$A = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1\beta} \\ a_{21} & a_{22} & \dots & a_{2\beta} \\ \vdots & \vdots & \ddots & \vdots \\ a_{\alpha 1} & a_{\alpha 2} & \dots & a_{\alpha\beta} \end{pmatrix}, \quad dA = \begin{pmatrix} da_{11} & da_{12} & \dots & da_{1\beta} \\ da_{21} & da_{22} & \dots & da_{2\beta} \\ \vdots & \vdots & \ddots & \vdots \\ da_{\alpha 1} & da_{\alpha 2} & \dots & da_{\alpha\beta} \end{pmatrix}$$

$$B = \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1\gamma} \\ b_{21} & b_{22} & \dots & b_{2\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ b_{\beta 1} & b_{\beta 2} & \dots & b_{\beta\gamma} \end{pmatrix}, \quad dB = \begin{pmatrix} db_{11} & db_{12} & \dots & db_{1\gamma} \\ db_{21} & db_{22} & \dots & db_{2\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ db_{\beta 1} & db_{\beta 2} & \dots & db_{\beta\gamma} \end{pmatrix}$$

$$C = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1\gamma} \\ c_{21} & c_{22} & \dots & c_{2\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ c_{\alpha 1} & c_{\alpha 2} & \dots & c_{\alpha\gamma} \end{pmatrix}, \quad dC = \begin{pmatrix} dc_{11} & dc_{12} & \dots & dc_{1\gamma} \\ dc_{21} & dc_{22} & \dots & dc_{2\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ dc_{\alpha 1} & dc_{\alpha 2} & \dots & dc_{\alpha\gamma} \end{pmatrix}$$

Matrix Multiplication and Expanded Jacobians

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1\beta} \\ a_{21} & a_{22} & \dots & a_{2\beta} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{i\beta} \\ \vdots & \vdots & & \vdots \\ a_{\alpha 1} & a_{\alpha 2} & \dots & a_{\alpha\beta} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1j} & \dots & b_{1\gamma} \\ b_{21} & b_{22} & \dots & b_{2j} & \dots & b_{2\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{\beta 1} & b_{\beta 2} & \dots & b_{\beta j} & \dots & b_{\beta\gamma} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1j} & \dots & c_{1\gamma} \\ c_{21} & c_{22} & \dots & c_{2j} & \dots & c_{2\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{i1} & c_{i2} & \dots & c_{ij} & \dots & c_{i\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{\alpha 1} & c_{\alpha 2} & \dots & c_{\alpha j} & \dots & c_{\alpha\gamma} \end{pmatrix}$$

$$\begin{aligned} \frac{\partial J}{\partial a_{ij}} &= \frac{\partial J}{\partial c_{i1}}\frac{\partial c_{i1}}{\partial a_{ij}} + \frac{\partial J}{\partial c_{i2}}\frac{\partial c_{i2}}{\partial a_{ij}} + \dots + \frac{\partial J}{\partial c_{i\gamma}}\frac{\partial c_{i\gamma}}{\partial a_{ij}} \\ &= dc_{i1} \cdot b_{j1} + dc_{i2} \cdot b_{j2} + \dots + dc_{i\gamma} \cdot b_{j\gamma} \\ &= dc_{i1} \cdot (b_T)_{1j} + dc_{i2} \cdot (b_T)_{2j} + \dots + dc_{i\gamma} \cdot (b_T)_{\gamma j} \\ &= Row_i(dC) \cdot Col_j(B^T) \end{aligned}$$

$$da_{ij} = Row_i(dC) \cdot Col_j(B^T)$$

Matrix Multiplication and Expanded Jacobians

$$da_{ij} = Row_i(dC) \cdot Col_j(B^T)$$

$$dA = \begin{pmatrix} da_{11} & da_{12} & \dots & da_{1\beta} \\ da_{21} & da_{22} & \dots & da_{2\beta} \\ \vdots & \vdots & \ddots & \vdots \\ da_{\alpha 1} & da_{\alpha 2} & \dots & da_{\alpha\beta} \end{pmatrix} = \begin{pmatrix} Row_1(dC) \cdot Col_1(B^T) & Row_1(dC) \cdot Col_2(B^T) & \dots & Row_1(dC) \cdot Col_\beta(B^T) \\ Row_2(dC) \cdot Col_1(B^T) & Row_2(dC) \cdot Col_2(B^T) & \dots & Row_2(dC) \cdot Col_\beta(B^T) \\ \vdots & \vdots & \ddots & \vdots \\ Row_\alpha(dC) \cdot Col_1(B^T) & Row_\alpha(dC) \cdot Col_2(B^T) & \dots & Row_\alpha(dC) \cdot Col_\beta(B^T) \end{pmatrix}$$

$$dA = dC \cdot B^T$$

$$C = A \cdot B \quad \Longrightarrow \quad dA = dC \cdot B^T$$

Matrix Multiplication and Expanded Jacobians

$$\begin{pmatrix} a_{11} & a_{12} & \dots & a_{1\beta} \\ a_{21} & a_{22} & \dots & a_{2\beta} \\ \vdots & \vdots & & \vdots \\ a_{i1} & a_{i2} & \dots & a_{i\beta} \\ \vdots & \vdots & & \vdots \\ a_{\alpha 1} & a_{\alpha 2} & \dots & a_{\alpha\beta} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} & \dots & b_{1j} & \dots & b_{1\gamma} \\ b_{21} & b_{22} & \dots & b_{2j} & \dots & b_{2\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ b_{\beta 1} & b_{\beta 2} & \dots & b_{\beta j} & \dots & b_{\beta\gamma} \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1j} & \dots & c_{1\gamma} \\ c_{21} & c_{22} & \dots & c_{2j} & \dots & c_{2\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{i1} & c_{i2} & \dots & c_{ij} & \dots & c_{i\gamma} \\ \vdots & \vdots & & \vdots & & \vdots \\ c_{\alpha 1} & c_{\alpha 2} & \dots & c_{\alpha j} & \dots & c_{\alpha\gamma} \end{pmatrix}$$

$$\frac{\partial J}{\partial b_{ij}} = \frac{\partial J}{\partial c_{1j}}\frac{\partial c_{1j}}{\partial b_{ij}} + \frac{\partial J}{\partial c_{2j}}\frac{\partial c_{2j}}{\partial b_{ij}} + \dots + \frac{\partial J}{\partial c_{\alpha j}}\frac{\partial c_{\alpha j}}{\partial b_{ij}}$$

$$= dc_{1j} \cdot a_{1i} + dc_{2j} \cdot a_{2i} + \dots + dc_{\alpha j} \cdot a_{\alpha i}$$

$$= (a_T)_{i1} \cdot dc_{1j} + (a_T)_{i2} \cdot dc_{2j} + \dots + (a_T)_{i\alpha} \cdot dc_{\alpha j}$$

$$= Row_i(A^T) \cdot Col_j(dC)$$

$$db_{ij} = Row_i(A^T) \cdot Col_j(dC)$$

$$dB = A^T \cdot dC \qquad \frac{\partial J}{\partial B} = A^T \cdot \frac{\partial J}{\partial C}$$

$$a_{1i}, a_{2i}, \dots, a_{\alpha i} \sim b_{ij}$$

$$\downarrow$$

$$c_{1j}, c_{2j}, \dots, c_{\alpha j}$$

Matrix Multiplication and Expanded Jacobians

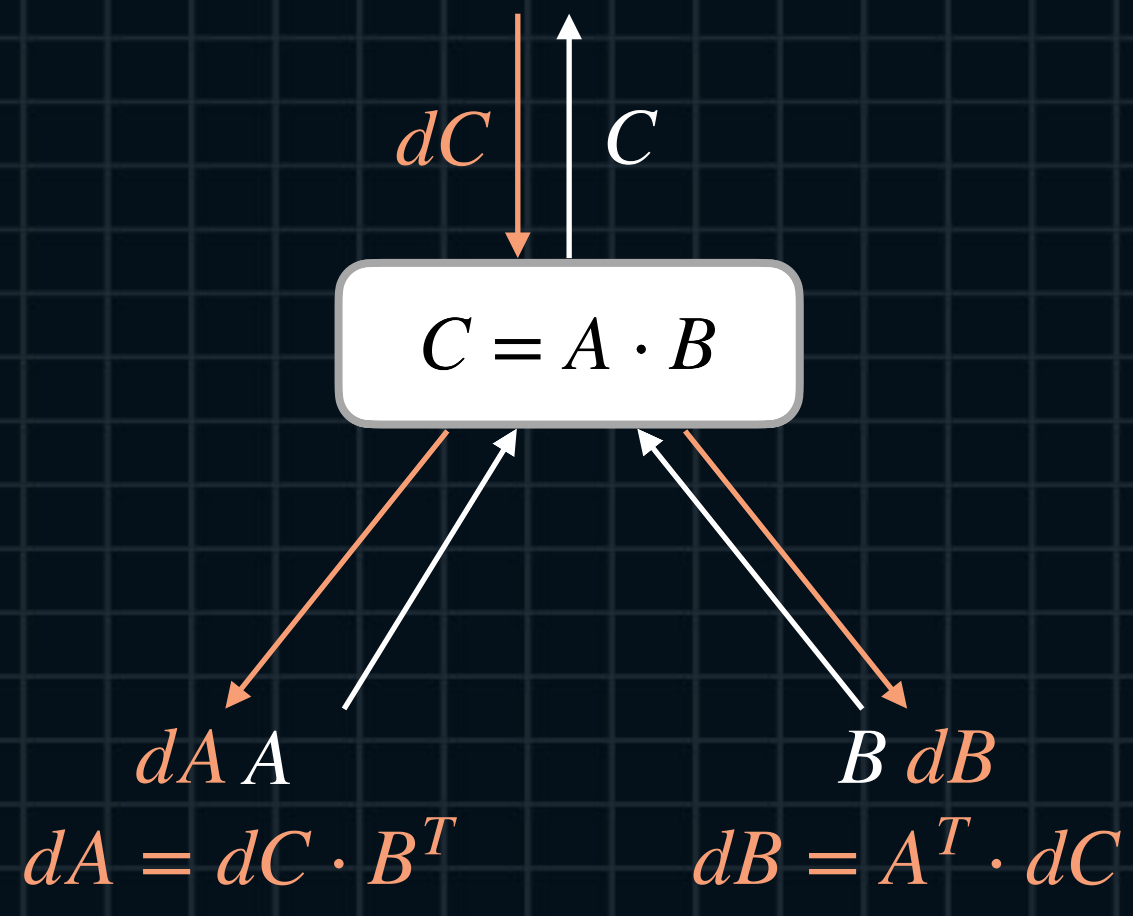
$$db_{ij} = Row_i(A^T) \cdot Col_j(dC)$$

$$dB = \begin{pmatrix} db_{11} & db_{12} & \dots & db_{1\gamma} \\ db_{21} & db_{22} & \dots & db_{2\gamma} \\ \vdots & \vdots & \ddots & \vdots \\ db_{\beta 1} & db_{\beta 2} & \dots & db_{\beta\gamma} \end{pmatrix} = \begin{pmatrix} Row_1(A^T) \cdot Col_1(dC) & Row_1(A^T) \cdot Col_2(dC) & \dots & Row_1(A^T) \cdot Col_\gamma(dC) \\ Row_2(A^T) \cdot Col_1(dC) & Row_2(A^T) \cdot Col_2(dC) & \dots & Row_2(A^T) \cdot Col_\gamma(dC) \\ \vdots & \vdots & \ddots & \vdots \\ Row_\beta(A^T) \cdot Col_1(dC) & Row_\beta(A^T) \cdot Col_2(dC) & \dots & Row_\beta(A^T) \cdot Col_\gamma(dC) \end{pmatrix}$$

$$dB = A^T \cdot dC$$

$$C = A \cdot B \quad \Longrightarrow \quad dB = A^T \cdot dC$$

Matrix Multiplication and Expanded Jacobians



Bias Vector and Expanded Jacobians

$$C = A + \left(\vec{b}\right)^T$$

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1\beta} \\ c_{21} & c_{22} & \dots & c_{2\beta} \\ \vdots & \vdots & \ddots & \vdots \\ c_{\alpha 1} & c_{\alpha 2} & \dots & c_{\alpha\beta} \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \dots & a_{1\beta} \\ a_{21} & a_{22} & \dots & a_{2\beta} \\ \vdots & \vdots & \ddots & \vdots \\ a_{\alpha 1} & a_{\alpha 2} & \dots & a_{\alpha\beta} \end{pmatrix} + \begin{pmatrix} b_1 & b_2 & \dots & b_\beta \end{pmatrix} = \begin{pmatrix} a_{11}+b_1 & a_{12}+b_2 & \dots & a_{1\beta}+b_\beta \\ a_{21}+b_1 & a_{22}+b_2 & \dots & a_{2\beta}+b_\beta \\ \vdots & \vdots & \ddots & \vdots \\ a_{\alpha 1}+b_1 & a_{\alpha 2}+b_2 & \dots & a_{\alpha\beta}+b_\beta \end{pmatrix}$$

$$\begin{pmatrix} \boxed{b_1} & \boxed{b_2} & \dots & \boxed{b_\beta} \end{pmatrix}$$

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1\beta} \\ c_{21} & c_{22} & \dots & c_{2\beta} \\ \vdots & \vdots & \ddots & \vdots \\ c_{\alpha 1} & c_{\alpha 2} & \dots & c_{\alpha\beta} \end{pmatrix}$$

Bias Vector and Expanded Jacobians

$$C = A + \left(\overrightarrow{b}\right)^T$$

$$\begin{pmatrix} c_{11} & c_{12} & \dots & c_{1\beta} \\ c_{21} & c_{22} & \dots & c_{2\beta} \\ \vdots & \vdots & \ddots & \vdots \\ c_{\alpha 1} & c_{\alpha 2} & \dots & c_{\alpha\beta} \end{pmatrix} = \begin{pmatrix} a_{11} + b_1 & a_{12} + b_2 & \dots & a_{1\beta} + b_\beta \\ a_{21} + b_1 & a_{22} + b_2 & \dots & a_{2\beta} + b_\beta \\ \vdots & \vdots & \ddots & \vdots \\ a_{\alpha 1} + b_1 & a_{\alpha 2} + b_2 & \dots & a_{\alpha\beta} + b_\beta \end{pmatrix}$$

$$\frac{\partial J}{\partial b_i} = \frac{\partial J}{\partial c_{1i}} \frac{\partial c_{1i}}{\partial b_i} + \frac{\partial J}{\partial c_{2i}} \frac{\partial c_{2i}}{\partial b_i} + \dots + \frac{\partial J}{\partial c_{\alpha i}} \frac{\partial c_{\alpha i}}{\partial b_i}$$

$$= dc_{1i} + dc_{2i} + \dots + dc_{\alpha i}$$

$$= \sum_{p=1}^{\alpha} dc_{pi}$$

Bias Vector and Expanded Jacobians

$$\frac{\partial J}{\partial b_i} = \sum_{p=1}^{\alpha} dc_{pi}$$

$$d\vec{b} = \begin{pmatrix} db_1 \\ db_2 \\ \vdots \\ db_\beta \end{pmatrix} = \begin{pmatrix} \sum_{p=1}^{\alpha} dc_{p1} \\ \sum_{p=1}^{\alpha} dc_{p2} \\ \vdots \\ \sum_{p=1}^{\alpha} dc_{p\beta} \end{pmatrix}$$

$$\begin{array}{cccc} \begin{pmatrix} \boxed{\begin{matrix} dc_{11} \\ dc_{21} \\ \vdots \\ dc_{\alpha 1} \end{matrix}} & \boxed{\begin{matrix} dc_{12} \\ dc_{22} \\ \vdots \\ dc_{\alpha 2} \end{matrix}} & \begin{matrix} \dots \\ \dots \\ \ddots \\ \dots \end{matrix} & \boxed{\begin{matrix} dc_{1\beta} \\ dc_{2\beta} \\ \vdots \\ dc_{\alpha\beta} \end{matrix}} \end{pmatrix} \\ \sum_{k=1}^{\alpha} dc_{k1} \quad \sum_{k=1}^{\alpha} dc_{k2} \qquad \sum_{k=1}^{\alpha} dc_{k\beta} \end{array}$$

$$d\vec{b} = sum(dC, axis = 0)$$

Affine Function and Expanded Jacobians

$$Z \in \mathbb{R}^{N \times l_O}$$

$$\uparrow$$

$$\boxed{Z = X \cdot W + \left(\vec{b}\right)^T} \qquad W \in \mathbb{R}^{l_I \times l_O},\ \vec{b} \in \mathbb{R}^{l_O}$$

$$\uparrow$$

$$X \in \mathbb{R}^{N \times l_I}$$

$$Z = X \cdot W + \left(\vec{b}\right)^T$$

$$(N, l_O) = (N, l_I) \cdot (l_I, l_O) + (1, l_O)$$

$$dX = dZ \cdot W^T$$

$$dW = X^T \cdot dZ \qquad d\vec{b} = sum\left(dZ, axis = 0\right)$$

Backpropagation and Jacobian Matrices

Lecture.10
Expanded Jacobians in Deep Learning