目录

矩阵、向量求导法则

复杂矩阵问题求导方法：可以从小到大，从scalar到vector再到matrix。

$$J(U,V) = \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2}(\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2)$$

$$= \sum_{i,j}\left(\sum_a U_{ia}V_{ja} - Y_{ij}\right)^2 + \frac{\lambda}{2}\left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2\right)$$

$$\frac{\partial J}{\partial U_{ia}} = \quad \cdots\cdots$$

$$= 2\sum_j \left(U_i V_j^T - Y_{ij}\right) V_{ja} + \lambda U_{ia},$$

$$= 2\left(U_i V^T - Y_i\right) V_{\cdot a} + \lambda U_{ia},$$

261229094735664.png

$$J(U,V) = \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2}(\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2)$$

$$= \sum_{i,j}\left(\sum_a U_{ia}V_{ja} - Y_{ij}\right)^2 + \frac{\lambda}{2}\left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2\right)$$

$$\frac{\partial J}{\partial U_{ia}} = 2\left(U_i V^T - Y_i\right) V_{\cdot a} + \lambda U_{ia}$$

$$\frac{\partial J}{\partial U_i} = 2\left(U_i V^T - Y_i\right) V + \lambda U_i$$

$$\frac{\partial J}{\partial U} = 2\left(UV^T - Y\right) V + \lambda U.$$

261224001048648.png

x is a column vector, A is a matrix

$$d(A * x)/dx = A$$

$$d(x^T * A)/dx^T = A$$

$$d(x^T * A)/dx = A^T$$

$$d(x^T * A * x)/dx = x^T(A^T + A)$$

enter description here

practice:

$$\frac{\partial \mathbf{a}^T \mathbf{X} \mathbf{b}}{\partial \mathbf{X}} = \mathbf{a}\mathbf{b}^T$$

$$\frac{\partial \mathbf{a}^T \mathbf{X}^T \mathbf{b}}{\partial \mathbf{X}} = \mathbf{b}\mathbf{a}^T$$

261238458939191.png

矩阵、向量求导法则

(1) 行向量对元素求导

设 $\mathbf{y}^T=\begin{bmatrix} y_1 & \cdots & y_n \end{bmatrix}$ 是 n 维行向量，x 是元素，则 $\dfrac{\partial \mathbf{y}^T}{\partial x}=\begin{bmatrix} \dfrac{\partial y_1}{\partial x} & \cdots & \dfrac{\partial y_n}{\partial x} \end{bmatrix}$。

(2) 列向量对元素求导

设 $\mathbf{y}=\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ 是 m 维列向量，x 是元素，则 $\dfrac{\partial \mathbf{y}}{\partial x}=\begin{bmatrix} \dfrac{\partial y_1}{\partial x} \\ \vdots \\ \dfrac{\partial y_m}{\partial x} \end{bmatrix}$。

(3) 矩阵对元素求导

设 $Y=\begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix}$ 是 $m\times n$ 矩阵，x 是元素，则

$$\frac{\partial Y}{\partial x}=\begin{bmatrix} \dfrac{\partial y_{11}}{\partial x} & \cdots & \dfrac{\partial y_{1n}}{\partial x} \\ \vdots & & \\ \dfrac{\partial y_{m1}}{\partial x} & \cdots & \dfrac{\partial y_{mn}}{\partial x} \end{bmatrix}。$$

(4) 元素对行向量求导

设 y 是元素，$\mathbf{x}^T=[x_1 \ \cdots \ x_q]$ 是 q 维行向量，则 $\dfrac{\partial y}{\partial \mathbf{x}^T}=\begin{bmatrix} \dfrac{\partial y}{\partial x_1} & \cdots & \dfrac{\partial y}{\partial x_q} \end{bmatrix}$。

(5) 元素对列向量求导

设 y 是元素，$\mathbf{x}=\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ 是 p 维列向量，则 $\dfrac{\partial y}{\partial \mathbf{x}}=\begin{bmatrix} \dfrac{\partial y}{\partial x_1} \\ \vdots \\ \dfrac{\partial y}{\partial x_p} \end{bmatrix}$。

(6) 元素对矩阵求导

设 y 是元素，$X=\begin{bmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & & \vdots \\ x_{p1} & \cdots & y_{pq} \end{bmatrix}$ 是 $p\times q$ 矩阵，则

$$\frac{\partial y}{\partial X}=\begin{bmatrix} \dfrac{\partial y}{\partial x_{11}} & \cdots & \dfrac{\partial y}{\partial x_{1q}} \\ \vdots & & \\ \dfrac{\partial y}{\partial x_{p1}} & \cdots & \dfrac{\partial y}{\partial x_{pq}} \end{bmatrix}。$$

261219153897663.gif

(7) 行向量对列向量求导

设 $\mathbf{y}^T = [y_1 \quad \cdots \quad y_n]$ 是 n 维行向量，$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ 是 p 维列向量，则

$$\frac{\partial \mathbf{y}^T}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_1} \\ \vdots & & \\ \frac{\partial y_1}{\partial x_p} & \cdots & \frac{\partial y_n}{\partial x_p} \end{bmatrix}。$$

(8) 列向量对行向量求导

设 $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ 是 m 维列向量，$\mathbf{x}^T = [x_1 \quad \cdots \quad x_q]$ 是 q 维行向量，则

$$\frac{\partial \mathbf{y}}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_q} \\ \vdots & & \\ \frac{\partial y_m}{\partial x_1} & \cdots & \frac{\partial y_m}{\partial x_q} \end{bmatrix}。$$

(9) 行向量对行向量求导

设 $\mathbf{y}^T = [y_1 \quad \cdots \quad y_n]$ 是 n 维行向量，$\mathbf{x}^T = [x_1 \quad \cdots \quad x_q]$ 是 q 维行向量，则

$$\frac{\partial \mathbf{y}^T}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial \mathbf{y}^T}{\partial x_1} & \cdots & \frac{\partial \mathbf{y}^T}{\partial x_q} \end{bmatrix}。$$

(10) 列向量对列向量求导

设 $\mathbf{y} = \begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ 是 m 维列向量，$\mathbf{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ 是 p 维列向量，则 $\frac{\partial \mathbf{y}}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial y_1}{\partial \mathbf{x}} \\ \vdots \\ \frac{\partial y_m}{\partial \mathbf{x}} \end{bmatrix}$。

(11) 矩阵对行向量求导

设 $Y = \begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix}$ 是 $m \times n$ 矩阵，$\mathbf{x}^T = [x_1 \quad \cdots \quad x_q]$ 是 q 维行向量，则

$$\frac{\partial Y}{\partial \mathbf{x}^T} = \begin{bmatrix} \frac{\partial Y}{\partial x_1} & \cdots & \frac{\partial Y}{\partial x_q} \end{bmatrix}。$$

261219289804718.gif

（12）矩阵对列向量求导

设 $Y=\begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix}$ 是 $m\times n$ 矩阵，$\mathbf{x}=\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ 是 p 维列向量，则

$$\frac{\partial Y}{\partial \mathbf{x}}=\begin{bmatrix} \frac{\partial y_{11}}{\partial \mathbf{x}} & \cdots & \frac{\partial y_{1n}}{\partial \mathbf{x}} \\ \vdots & & \vdots \\ \frac{\partial y_{m1}}{\partial \mathbf{x}} & \cdots & \frac{\partial y_{mn}}{\partial \mathbf{x}} \end{bmatrix}。$$

（13）行向量对矩阵求导

设 $\mathbf{y}^T=[y_1 \quad \cdots \quad y_n]$ 是 n 维行向量，$X=\begin{bmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & & \vdots \\ x_{p1} & \cdots & y_{pq} \end{bmatrix}$ 是 $p\times q$ 矩阵，则

$$\frac{\partial \mathbf{y}^T}{\partial X}=\begin{bmatrix} \frac{\partial \mathbf{y}^T}{\partial x_{11}} & \cdots & \frac{\partial \mathbf{y}^T}{\partial x_{1q}} \\ \vdots & & \\ \frac{\partial \mathbf{y}^T}{\partial x_{p1}} & \cdots & \frac{\partial \mathbf{y}^T}{\partial x_{pq}} \end{bmatrix}。$$

（14）列向量对矩阵求导

设 $\mathbf{y}=\begin{bmatrix} y_1 \\ \vdots \\ y_m \end{bmatrix}$ 是 m 维列向量，$X=\begin{bmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & & \vdots \\ x_{p1} & \cdots & y_{pq} \end{bmatrix}$ 是 $p\times q$ 矩阵，则

$$\frac{\partial \mathbf{y}}{\partial X}=\begin{bmatrix} \frac{\partial y_1}{\partial X} \\ \vdots \\ \frac{\partial y_m}{\partial X} \end{bmatrix}。$$

（15）矩阵对矩阵求导

设 $Y=\begin{bmatrix} y_{11} & \cdots & y_{1n} \\ \vdots & & \vdots \\ y_{m1} & \cdots & y_{mn} \end{bmatrix}=\begin{bmatrix} \mathbf{y}_1^T \\ \vdots \\ \mathbf{y}_m^T \end{bmatrix}$ 是 $m\times n$ 矩阵，$X=\begin{bmatrix} x_{11} & \cdots & x_{1q} \\ \vdots & & \vdots \\ x_{p1} & \cdots & y_{pq} \end{bmatrix}$

$=[\mathbf{x}_1 \quad \cdots \quad \mathbf{x}_q]$ 是 $p\times q$ 矩阵，则

$$\frac{\partial Y}{\partial X}=\begin{bmatrix} \frac{\partial Y}{\partial \mathbf{x}_1} & \cdots & \frac{\partial Y}{\partial \mathbf{x}_q} \end{bmatrix}=\begin{bmatrix} \frac{\partial \mathbf{y}_1^T}{\partial X} \\ \vdots \\ \frac{\partial \mathbf{y}_m^T}{\partial X} \end{bmatrix}=\begin{bmatrix} \frac{\partial \mathbf{y}_1^T}{\partial \mathbf{x}_1} & \cdots & \frac{\partial \mathbf{y}_1^T}{\partial \mathbf{x}_q} \\ \vdots & & \vdots \\ \frac{\partial \mathbf{y}_m^T}{\partial \mathbf{x}_1} & \cdots & \frac{\partial \mathbf{y}_m^T}{\partial \mathbf{x}_q} \end{bmatrix}。$$

261219434733214.gif

例 设 $\frac{\partial A}{\partial X}=\begin{bmatrix} 2xy & y^2 & y \\ x^2 & 2xy & x \end{bmatrix}$，$X=\begin{bmatrix} x \\ y \end{bmatrix}$，根据（12）矩阵对列向量求导法则，有

$$\frac{\partial^2 A}{\partial X^2}=\begin{bmatrix} \frac{\partial(2xy)}{\partial X} & \frac{\partial(y^2)}{\partial X} & \frac{\partial y}{\partial X} \\ \frac{\partial(x^2)}{\partial X} & \frac{\partial(2xy)}{\partial X} & \frac{\partial x}{\partial X} \end{bmatrix}=\begin{bmatrix} 2y & 0 & 0 \\ 2x & 2y & 1 \\ 2x & 2y & 1 \\ 0 & 2x & 0 \end{bmatrix}。$$

例 设 $Y=\begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix}$，$X=\begin{bmatrix} u & x \\ v & y \\ w & z \end{bmatrix}$，根据（15）矩阵对矩阵求导法则，有

$$\frac{\partial Y}{\partial X}=\begin{bmatrix} \frac{\partial\begin{bmatrix} a & b & c \end{bmatrix}}{\partial\begin{bmatrix} u \\ v \\ w \end{bmatrix}} & \frac{\partial\begin{bmatrix} a & b & c \end{bmatrix}}{\partial\begin{bmatrix} x \\ y \\ z \end{bmatrix}} \\ \frac{\partial\begin{bmatrix} d & e & f \end{bmatrix}}{\partial\begin{bmatrix} u \\ v \\ w \end{bmatrix}} & \frac{\partial\begin{bmatrix} d & e & f \end{bmatrix}}{\partial\begin{bmatrix} x \\ y \\ z \end{bmatrix}} \end{bmatrix}=\begin{bmatrix} \frac{\partial a}{\partial u} & \frac{\partial b}{\partial u} & \frac{\partial c}{\partial u} & \frac{\partial a}{\partial x} & \frac{\partial b}{\partial x} & \frac{\partial c}{\partial x} \\ \frac{\partial a}{\partial v} & \frac{\partial b}{\partial v} & \frac{\partial c}{\partial v} & \frac{\partial a}{\partial y} & \frac{\partial b}{\partial y} & \frac{\partial c}{\partial y} \\ \frac{\partial a}{\partial w} & \frac{\partial b}{\partial w} & \frac{\partial c}{\partial w} & \frac{\partial a}{\partial z} & \frac{\partial b}{\partial z} & \frac{\partial c}{\partial z} \\ \frac{\partial d}{\partial u} & \frac{\partial e}{\partial u} & \frac{\partial f}{\partial u} & \frac{\partial d}{\partial x} & \frac{\partial e}{\partial x} & \frac{\partial f}{\partial x} \\ \frac{\partial d}{\partial v} & \frac{\partial e}{\partial v} & \frac{\partial f}{\partial v} & \frac{\partial d}{\partial y} & \frac{\partial e}{\partial y} & \frac{\partial f}{\partial y} \\ \frac{\partial d}{\partial w} & \frac{\partial e}{\partial w} & \frac{\partial f}{\partial w} & \frac{\partial d}{\partial z} & \frac{\partial e}{\partial z} & \frac{\partial f}{\partial z} \end{bmatrix}。$$