Matrix Calculus

Steve Canves

yqykrhf@163.com

1 Scalar With Respect To Matrix

Think of the defination of derivative of the scalar f with respect to matrix $X_{m\times n}$

$$\frac{\partial f}{\partial X} = \left[\frac{\partial f}{\partial X_{ij}}\right]_{m\times n} \tag{1}$$

i.e. derivative of f with respect to X's by element, then which is arranged into a matrix with the same size to X. However, derivation by element breaks the integrity. Let's find ahother approach from whole aspect.

Review derivative in one variable(scalar with respect to scalar):

$$df = f'(x)dx$$

And derivative in mutiple variables(scalar with respect to vector) is

$$df = \sum_{i=1}^{n}\frac{\partial f}{\partial x_i}dx_i = (\frac{\partial f}{\partial \boldsymbol{x}})^T d\boldsymbol{x}$$

Similiarily, relate matrix derivative to differentiation:

$$df = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{\partial f}{\partial X_{ij}}dX_{ij} = tr((\frac{\partial f}{\partial X})^T dX) \tag{2}$$

Here X is matrix, X_{ij} is element. If X is a vector, then $df = (\frac{\partial f}{\partial X})^T dX$

Note: $tr(A^TB) = \sum_{i,j}A_{ij}B_{ij}$

Rules for matrix differentiation:

1. $d(X \pm Y) = dX \pm dY;\ d(XY) = d(X)Y + Xd(Y);\ d(X^T) = (dX)^T;\ dtr(X) = tr(dX)$
2. $dX^{-1} = -x^{-1}dXX^{-1}$. Both sides differentiate $XX^{-1} = I$ to prove it.
3. $d(|X|) = tr(X^{adjoint}dX) = |X|tr(X^{-1}dX)$
4. element-wise mutiplication:

$$d(X \odot Y) = dX \odot Y + X \odot dY$$

5. element-wise function:

$$d\sigma(X) = \sigma'(X) \odot dX$$

$$X = \begin{bmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \end{bmatrix}, d\sin(X) = \begin{bmatrix} \cos X_{11}dX_{11} & \cos X_{12}dX_{12} \\ \cos X_{21}dX_{21} & \cos X_{22}dX_{22} \end{bmatrix} = \cos(X) \odot dX$$

Trace trick:

1. $tr(scalar) = a$
2. $tr(A^T) = tr(A)$
3. $tr(A \pm B) = tr(A) \pm tr(B)$
4. $tr(AB) = tr(BA) = \sum_{i,j} A_{ij} B_{ji}$
5. $tr(A^T(B \odot C)) = tr((A \odot B)^T C) = \sum_{i,j} A_{ij} B_{ij} C_{ij}$

Examples:

ex1.

$f = \boldsymbol{a}^T X \boldsymbol{b}$, solve $\frac{\partial f}{\partial X}$. f is scalar, **a** is $m \times 1$ column vector, X is $m \times n$ matrix, **b** is $n \times 1$

$$df = da^T Xb + a^T dXb + a^T Xdb = a^T dXb$$

then trace both sides:

$$df = tr(a^T dXb) = tr(ba^T dX) = tr((ab^T)^T dX)$$

then

$$\frac{\partial f}{\partial X} = ab^T$$

ex2.

$f = \boldsymbol{a}^T exp(X\boldsymbol{b})$, solve $\frac{\partial f}{\partial X}$. f is scalar, **a** is $m \times 1$ column vector, X is $m \times n$ matrix, **b** is $n \times 1$ exp is element-wise operation.

$$df = a^T(exp(Xb) \odot d(Xb))$$

then trace both sides:

$$\begin{aligned} df &= tr((a \odot exp(Xb))^T dXb) \\ &= tr(b(a \odot exp(Xb))^T dX) \\ &= tr(((a \odot exp(Xb))b^T)^T dX) \end{aligned}$$

then

$$\frac{\partial f}{\partial X} = (a \odot exp(Xb))b^T$$

ex3.

$f = tr(Y^T MY)$,$Y = \sigma(WX)$, solve $\frac{\partial f}{\partial X}$. f is scalar, **W** is $l \times m$ matrix, X is $m \times n$ matrix, Y is $l \times n$ matrix, M is $l \times l$ symetric matrix, σ is element-wise operation.

$$\begin{aligned} df &= tr(d(Y)^T)MY) + tr(Y^T MdY) \\ &= tr(Y^T M^T dY) + tr(Y^T MdY) \\ &= tr(Y^T(M + M^T)dY) \\ &= tr(Y^T 2MdY) \end{aligned}$$

$\frac{\partial f}{\partial Y} = 2M^TY = 2MY$

$$
\begin{aligned}
df &= tr((\frac{\partial f}{\partial Y})^T dY) \\
&= tr((\frac{\partial f}{\partial Y})^T (\sigma'(WX) \odot d(WX))) \\
&= tr\left((\frac{\partial f}{\partial Y})^T (\sigma'(WX) \odot WdX)\right) \\
&= tr\left((\frac{\partial f}{\partial Y} \odot \sigma'(WX))^T WdX\right) \\
&= tr\left((W^T(\frac{\partial f}{\partial Y} \odot \sigma'(WX)))^T dX\right)
\end{aligned}
$$

$\frac{\partial f}{\partial X} = W^T(2MY \odot \sigma'(WX))$

ex4. least square estimation

$E = ||AX - b||^2$, solve $\frac{\partial E}{\partial X}$'s zero point. E is scalar, A is $m \times n$ matrix, X is $n \times 1$ column vector, b is $m \times 1$ column vector.

$$E = (AX - b)^T(AX - b)$$

$$
\begin{aligned}
dE &= (AX - b)^T(AX - b) \\
&= (AdX)^T(AX - b) + (AX - b)^T(AdX) \\
&= 2(AX - b)^T(AdX)
\end{aligned}
$$

$$\frac{\partial E}{\partial X} = 2A^T(AX - b) = 0$$

then

$$A^TAX = A^Tb$$

2 Matrix With Respect To Matrix

Define derivative of vector with respect to vector firstly. Derivative of vector $\boldsymbol{f}_{p\times 1}$ with respect to vector $\boldsymbol{x}$ is defined as:

$$\frac{\partial \boldsymbol{f}}{\partial \boldsymbol{x}} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_2}{\partial x_1} & \cdots & \frac{\partial f_p}{\partial x_1} \\ \frac{\partial f_1}{\partial x_2} & \frac{\partial f_2}{\partial x_2} & \cdots & \frac{\partial f_p}{\partial x_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial f_1}{\partial x_m} & \frac{\partial f_2}{\partial x_m} & \cdots & \frac{\partial f_p}{\partial x_m} \end{bmatrix} \tag{3}$$

3 Reference

- This blog.
- This blog.