Matrix Calculus and Algebra

Yi Zhang

Outline

- Matrix calculus and algebra
  - Dimensions of derivatives
  - Basic calculations of matrix derivatives
  - Rules for product, chain, trace, determinant and norms

Matrix Derivatives: Dimensions

- The basic case: y scalar, x scalar
  - dy/dx: scalar

- Generalize to vectors/matrices

Matrix Derivatives: Dimensions [T. Minka]

	Scalar	Vector	Matrix
Scalar	$\frac{\mathrm{d}y}{\mathrm{d}x}$	$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}x} = \left[\frac{\partial y_i}{\partial x}\right]$	$\frac{\mathrm{d}\mathbf{Y}}{\mathrm{d}x} = \left[\frac{\partial y_{ij}}{\partial x}\right]$
Vector	$\frac{\mathrm{d}y}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial y}{\partial x_j}\right]$	$\frac{\mathrm{d}\mathbf{y}}{\mathrm{d}\mathbf{x}} = \left[\frac{\partial y_i}{\partial x_j}\right]$	
Matrix	$\frac{\mathrm{d}y}{\mathrm{d}\mathbf{X}} = \left[\frac{\partial y}{\partial x_{ji}}\right]$		

Basic Calculations

- Step 1: know the dimensions
- Step 2: element-wise calculations
- Step 3: put into the vector/matrix form

Basic Calculations

- Example:

$$\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}}$$

  - Dimensions? Element-wise? Vector/matrix form?

Basic Calculations

- Example:

$$\frac{\partial \mathbf{a}^T \mathbf{x}}{\partial \mathbf{x}} = \mathbf{a}$$

Basic Calculations

- Too simple? Try this … [J. Rennie]

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

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

$$U \in \mathbb{R}^{n\times k},\ V \in \mathbb{R}^{m\times k},\ \text{and } Y \in \mathbb{R}^{n\times m}$$

- We want: $\dfrac{\partial J}{\partial U}$

Basic Calculations

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

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

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

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

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

Basic Calculations

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

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

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

Basic Calculations

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

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

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

$$\frac{\partial J}{\partial U_i} =$$

Basic Calculations

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

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

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

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

Basic Calculations

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

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

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

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

$$\frac{\partial J}{\partial U} =$$

Basic Calculations

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

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

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

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

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

Some Examples

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

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

Gradient and Hessian

$$J(U,V) = \|UV^T - Y\|^2_{\text{Fro}} + \frac{\lambda}{2}(\|U\|^2_{\text{Fro}} + \|V\|^2_{\text{Fro}})$$
$$= \sum_{i,j}\left(\sum_a U_{ia}V_{ja} - Y_{ij}\right)^2 + \frac{\lambda}{2}\left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2\right)$$

$$f \quad = \quad \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x}$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} \quad =$$

Gradient and Hessian

$$J(U,V) = \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2}(\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2)$$
$$= \sum_{i,j}\left(\sum_a U_{ia}V_{ja} - Y_{ij}\right)^2 + \frac{\lambda}{2}\left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2\right)$$

$$f = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x}$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x} + \mathbf{b}$$

$$\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^T} =$$

Gradient and Hessian

$$J(U,V) = \|UV^T - Y\|_{\text{Fro}}^2 + \frac{\lambda}{2}(\|U\|_{\text{Fro}}^2 + \|V\|_{\text{Fro}}^2)$$
$$= \sum_{i,j}\left(\sum_a U_{ia}V_{ja} - Y_{ij}\right)^2 + \frac{\lambda}{2}\left(\sum_{i,a} U_{ia}^2 + \sum_{j,a} V_{ja}^2\right)$$

$$f = \mathbf{x}^T\mathbf{A}\mathbf{x} + \mathbf{b}^T\mathbf{x}$$

$$\nabla_{\mathbf{x}} f = \frac{\partial f}{\partial \mathbf{x}} = (\mathbf{A} + \mathbf{A}^T)\mathbf{x} + \mathbf{b}$$

$$\frac{\partial^2 f}{\partial \mathbf{x} \partial \mathbf{x}^T} = \mathbf{A} + \mathbf{A}^T$$

Some Basic Rules

- Take derivatives w.r.t. *z* (omitted)

$$
\begin{aligned}
\partial \mathbf{A} &= 0 \quad (\mathbf{A} \text{ is a constant}) \\
\partial (\mathbf{X} + \mathbf{Y}) &= \partial \mathbf{X} + \partial \mathbf{Y} \\
\partial (\alpha \mathbf{X}) &= \alpha \partial \mathbf{X}
\end{aligned}
$$

Product Rules

- Take derivatives w.r.t. *z* (omitted)

$$\begin{aligned}
\partial(\mathbf{X} \circ \mathbf{Y}) &= (\partial \mathbf{X}) \circ \mathbf{Y} + \mathbf{X} \circ (\partial \mathbf{Y}) \\
\partial(\mathbf{X}\mathbf{Y}) &= (\partial \mathbf{X})\mathbf{Y} + \mathbf{X}(\partial \mathbf{Y}) \\
\partial(\mathbf{X} \otimes \mathbf{Y}) &= (\partial \mathbf{X}) \otimes \mathbf{Y} + \mathbf{X} \otimes (\partial \mathbf{Y})
\end{aligned}$$

Derivatives of Determinants

- Take derivatives w.r.t. *z* (omitted)

$$\begin{aligned}
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) \\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X})
\end{aligned}$$

The Chain Rule

- Suppose $\mathbf{U} = f(\mathbf{X})$

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}}$$

- The chain rule:

$$\frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^{M} \sum_{l=1}^{N} \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}}$$

The Chain Rule

- Suppose $\mathbf{U} = f(\mathbf{X})$

$$\frac{\partial g(\mathbf{U})}{\partial \mathbf{X}} = \frac{\partial g(f(\mathbf{X}))}{\partial \mathbf{X}}$$

- The chain rule:

$$\frac{\partial g(\mathbf{U})}{\partial x_{ij}} = \sum_{k=1}^{M} \sum_{l=1}^{N} \frac{\partial g(\mathbf{U})}{\partial u_{kl}} \frac{\partial u_{kl}}{\partial x_{ij}}$$

$$\frac{\partial g(\mathbf{U})}{\partial X_{ij}} = \mathrm{Tr}\Big[(\frac{\partial g(\mathbf{U})}{\partial \mathbf{U}})^T \frac{\partial \mathbf{U}}{\partial X_{ij}} \Big]$$

Derivatives of Traces [K.B. Petersen]

- Assume *F*(***X***) is an element-wise differentiable function
  - f() is the scalar derivative of F()

$$\frac{\partial \mathrm{Tr}(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^T$$

Derivatives of Traces [K.B. Petersen]

- Given: $$\frac{\partial \mathrm{Tr}(F(\mathbf{X}))}{\partial \mathbf{X}} = f(\mathbf{X})^T$$

$$\begin{aligned}
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}\mathbf{A}) &= \mathbf{A}^T \\
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}\mathbf{B}) &= \mathbf{A}^T\mathbf{B}^T \\
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{A}\mathbf{X}^T\mathbf{B}) &= \mathbf{B}\mathbf{A} \\
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}^T\mathbf{A}) &= \mathbf{A} \\
\frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}^2) &= 2\mathbf{X}^T
\end{aligned}$$

Derivatives of Frobenius Norm

- Frobenius norm

$$||\mathbf{A}||_{\mathrm{F}} = \sqrt{\sum_{ij} |A_{ij}|^2} = \sqrt{\mathrm{Tr}(\mathbf{A}\mathbf{A}^H)}$$

- Derivatives

$$\frac{\partial}{\partial \mathbf{X}} ||\mathbf{X}||_{\mathrm{F}}^2 = 2\mathbf{X} = \frac{\partial}{\partial \mathbf{X}} \mathrm{Tr}(\mathbf{X}\mathbf{X}^H)$$

References

- T. Minka. Old and New Matrix Algebra Useful for Statistics
- K. B. Petersen. The Matrix Cookbook
- J. D. M. Rennie. A Simple Exercise on Matrix Derivatives.