THE UNIVERSITY of ADELAIDE

Overview of Machine Learning and Essential Mathematic Skills for Machine Learning

Dr Alfred Krzywicki based on Dr Lingqiao Liu's slides
University of Adelaide

Summation and Product

- Commonly used operations in Statistic Machine Learning
- Summation notations
  - Summation

$$\sum_{i=1}^{N} a_i = \sum_i a_i = \sum_j a_j$$

  - Summation with two indices

$$\sum_{i=1}^{M} \sum_{j=1}^{N} a_{ij} = \sum_{i,j} a_{ij}$$

Summation and Product

- A useful formula (a little bit counter-intuitive)

$$\sum_{i=1}^{M} \sum_{j=1}^{N} a_i b_j = \left(\sum_i a_i\right) \left(\sum_j b_j\right) = \left(\sum_i a_i\right) \left(\sum_i b_i\right)$$

Summation and Product

- A useful formula (a little bit counter-intuitive)

$$\sum_{i=1}^{M} \sum_{j=1}^{N} a_i b_j = \left(\sum_i a_i\right) \left(\sum_j b_j\right) = \left(\sum_i a_i\right) \left(\sum_i b_i\right)$$

- Proof

$$\sum_{i=1}^{M} \sum_{j=1}^{N} a_i b_j = \sum_i a_i \left(\sum_j b_j\right) = \left(\sum_i b_i\right) \left(\sum_i a_i\right)$$

Linear algebra: vector, matrix and basic matrix operations

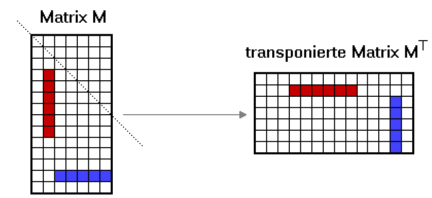
- Vectors and matrix

Scalar	Vector	Matrix
1	$\begin{bmatrix}1\\2\end{bmatrix}$	$\begin{bmatrix}1&2\\3&4\end{bmatrix}$

- Basic operations
  - Multiplication

$$\begin{bmatrix}a_1&a_2&a_3\\a_4&a_5&a_6\\a_7&a_8&a_9\end{bmatrix}\begin{bmatrix}b_1&b_2&b_3\\b_4&b_5&b_6\\b_7&b_8&b_9\end{bmatrix}=\begin{bmatrix}c_1&c_2&c_3\\c_4&c_5&c_6\\c_7&c_8&c_9\end{bmatrix}$$

  - Transpose



$$\begin{bmatrix}3&3\\5&-3\\1&3\\0&-2\\5&-2\end{bmatrix}^T=\begin{bmatrix}3&5&1&0&5\\3&-3&3&-2&-2\end{bmatrix}$$

  - Inverse

$$\mathbf{A}\mathbf{A}^{-1}=\mathbf{I}$$

Matrix multiplication

- View matrix as a set of vectors

$$\mathbf{A}\mathbf{b} = [\underbrace{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n}_{\text{Row vectors}}]^T \mathbf{b} = [\mathbf{a}_1^\top b, \mathbf{a}_2^\top b, \cdots, \mathbf{a}_n^\top b]^\top$$

$$\mathbf{A}\boldsymbol{\Lambda} = [\underbrace{\mathbf{a}_1, \mathbf{a}_2, \cdots, \mathbf{a}_n}_{\text{Column vectors}}] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} = [\lambda_1 \mathbf{a}_1, \lambda_2 \mathbf{a}_2, \cdots, \lambda_n \mathbf{a}_n]$$

Matrix product (example)

- Multiply row value by corresponding col value and add to make one value of the product, located at the intersection of row and col.
- Generally not commutative
- **x** **y** **z = x*y**
- [1 2 3] [1 **2** 3] [. . .]
- [**2** **2** **3**] * [2 **2** 3] = [. **17** .]
- [3 3 3] [3 **3** 3] [. . .]

- z22 = x21*y12 + x22*y22 + x23*y32

Inner product and norms

- Inner product between two vectors

$$< \mathbf{x}, \mathbf{y} >= \mathbf{x}^T \mathbf{y} = \sum_i x_i y_i$$

- Vector Norms
  - Measure the length of the vector
  - Not unique: could have infinite number of definitions
  - Commonly used ones

$$l_2 \text{ norm: } \|x\|_2 = \sqrt{\sum_i x_i^2} \qquad l_1 \text{ norm: } \|x\|_1 = \sum_i |x_i|$$

$$l_p \text{ norm: } \|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$$

Trace and Matrix Norm

- Definition $Tr(A) = \sum_i a_{ii} \quad Tr(a) = a$
- Properties

$$Tr(X^\top Y) = Tr(XY^\top) = Tr(Y^\top X) = Tr(YX^\top)$$

$$Tr(A) + Tr(B) = Tr(A + B)$$

- Frobenius norm

$$\|A\|_F = \sqrt{\sum_{ij} a_{ij}^2}$$

- Relationship to Trace

$$\|A\|_F = \sqrt{\sum_{ij} a_{ij}^2} = \sqrt{Tr(AA^\top)} = \sqrt{Tr(A^\top A)}$$

Matrix calculus

- For functions that involve matrices or vectors
  - Case 1: Vector/Matrix variable and scalar output
  - Case 2: Vector/Matrix variable and vector output
- Definition

$$\frac{\partial y}{\partial \mathbf{x}} \stackrel{\text{def}}{=} \begin{bmatrix} \frac{\partial y}{\partial x_1} \\ \frac{\partial y}{\partial x_2} \\ \vdots \\ \frac{\partial y}{\partial x_n} \end{bmatrix}.$$

- Application
  - Similar

Matrix calculus

- Properties
- More info
  - [Matrix Calculus]
- Trick to memorize
  - Analogy to scalar ca
  - Check dimensions

Identities: vector-by-vector $\frac{\partial \mathbf{y}}{\partial \mathbf{x}}$

Condition	Expression	Numerator layout, i.e. by $\mathbf{y}$ and $\mathbf{x}^T$	Denominator layout, i.e. by $\mathbf{y}^T$ and $\mathbf{x}$
$\mathbf{a}$ is not a function of $\mathbf{x}$	$\frac{\partial \mathbf{a}}{\partial \mathbf{x}} =$	$\mathbf{0}$	
	$\frac{\partial \mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{I}$	
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\frac{\partial \mathbf{A}\mathbf{x}}{\partial \mathbf{x}} =$	$\mathbf{A}$	$\mathbf{A}^\top$
$\mathbf{A}$ is not a function of $\mathbf{x}$	$\frac{\partial \mathbf{x}^\top \mathbf{A}}{\partial \mathbf{x}} =$	$\mathbf{A}^\top$	$\mathbf{A}$
a is not a function of $\mathbf{x}$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial a\mathbf{u}}{\partial \mathbf{x}} =$	$a\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	
$v = v(\mathbf{x})$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial v\mathbf{u}}{\partial \mathbf{x}} =$	$v\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \mathbf{u}\frac{\partial v}{\partial \mathbf{x}}$	$v\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial v}{\partial \mathbf{x}}\mathbf{u}^\top$
$\mathbf{A}$ is not a function of $\mathbf{x}$, $\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{A}\mathbf{u}}{\partial \mathbf{x}} =$	$\mathbf{A}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\mathbf{A}^\top$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$, $\mathbf{v} = \mathbf{v}(\mathbf{x})$	$\frac{\partial (\mathbf{u}+\mathbf{v})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}} + \frac{\partial \mathbf{v}}{\partial \mathbf{x}}$	
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}$
$\mathbf{u} = \mathbf{u}(\mathbf{x})$	$\frac{\partial \mathbf{f}(\mathbf{g}(\mathbf{u}))}{\partial \mathbf{x}} =$	$\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{u}}{\partial \mathbf{x}}$	$\frac{\partial \mathbf{u}}{\partial \mathbf{x}}\frac{\partial \mathbf{g}(\mathbf{u})}{\partial \mathbf{u}}\frac{\partial \mathbf{f}(\mathbf{g})}{\partial \mathbf{g}}$

Matrix calculus

- More information
- Exercise

$$\min_{\mathbf{x}} \|Ax - b\|_2^2$$

- Hint

$$\|Ax - b\|_2^2 = (Ax - b)^\top (Ax - b)$$

Matrix calculus

- More information
- Exercise

$$\min_{\mathbf{x}} \|Ax - b\|_2^2$$

- Hint

$$\|Ax - b\|_2^2 = (Ax - b)^\top (Ax - b)$$

$$\frac{\partial \|Ax-b\|_2^2}{\partial x} = 2A^\top (Ax - b) = 2A^\top Ax - 2A^\top b$$

$$Solve \quad \frac{\partial \|Ax-b\|_2^2}{\partial x} = 2A^\top Ax - 2A^\top b = 0$$

$$x = (A^\top A)^{-1} A^\top b$$