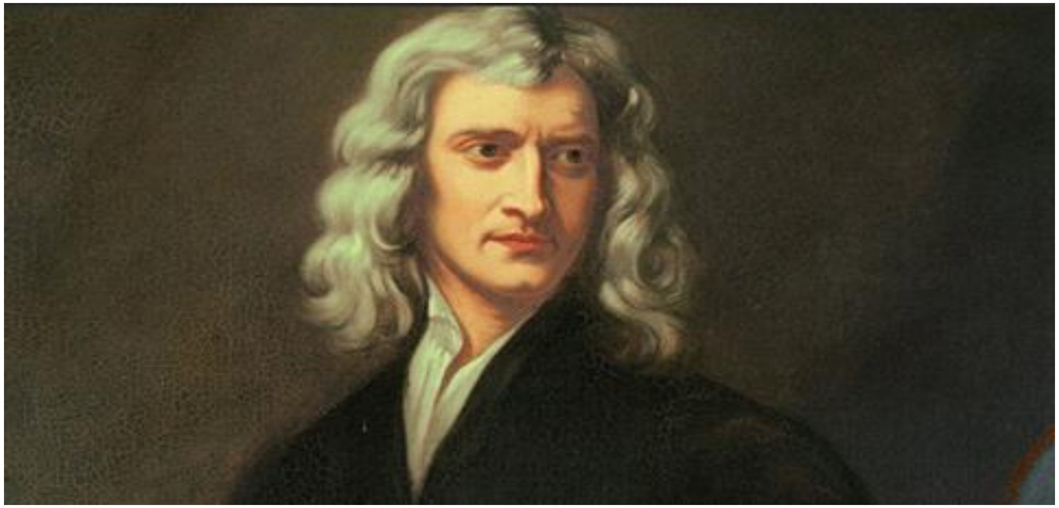
# 机器学习数学基础

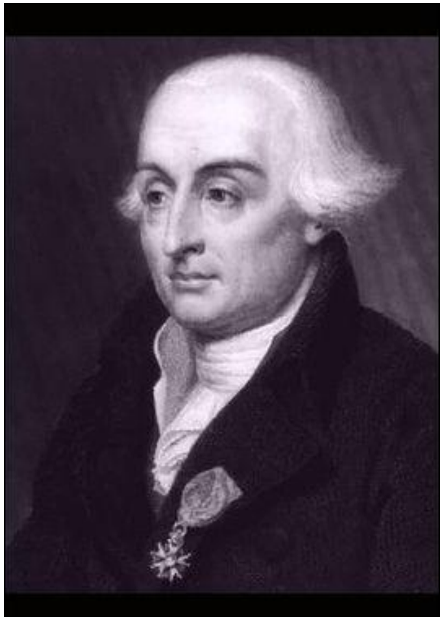
讲师：武晟然

# 主要内容

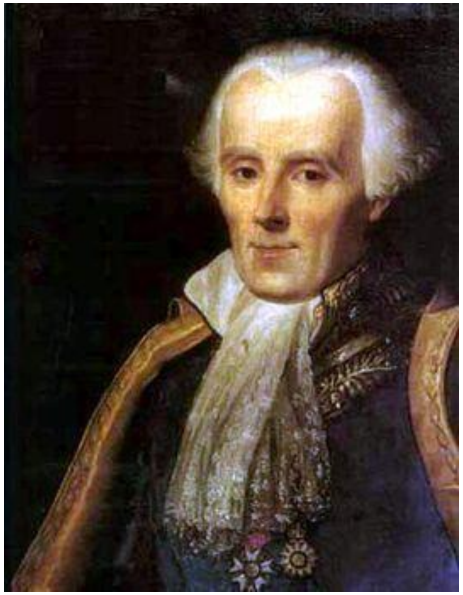
- 线性代数知识
- 微积分知识
- 概率与统计知识



牛顿



拉格朗日



拉普拉斯

高斯

牛顿

拉格朗日

拉普拉斯

高斯

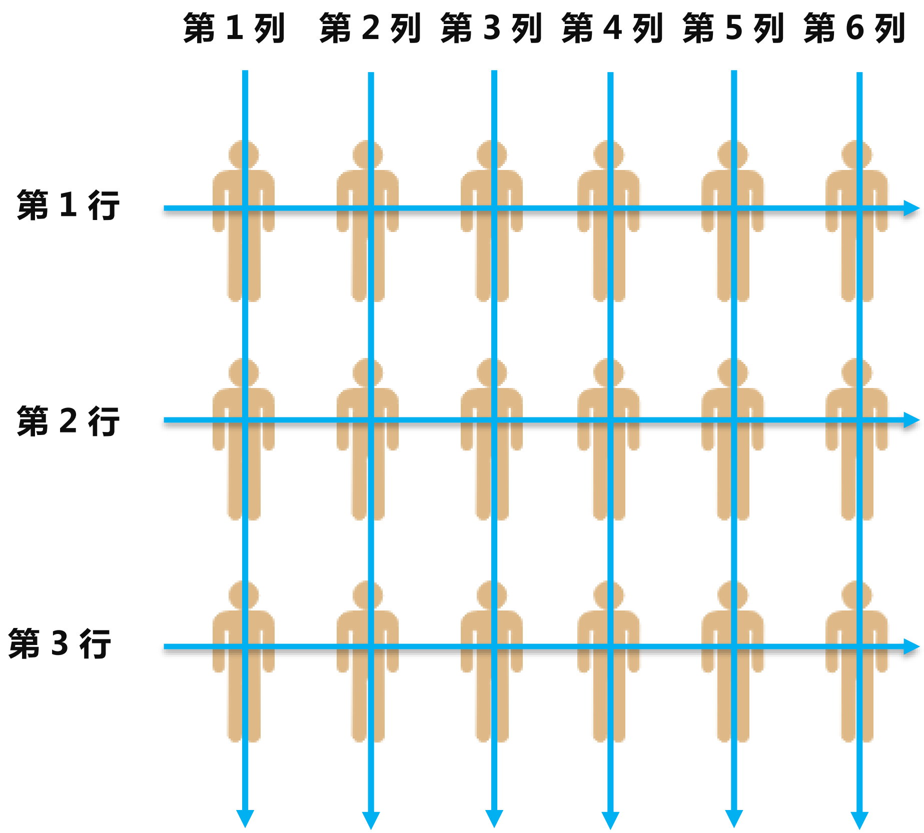
# 线性代数

- 什么是矩阵
- 矩阵中的基本概念
- 矩阵的加法
- 矩阵的乘法
- 矩阵的转置
- 矩阵的运算法则
- 矩阵的逆

# 矩阵

- 矩阵（Matrix）是一个按照长方阵列排列的复数或实数集合
- 矩阵最早来自于方程组的系数及常数所构成的方阵，最初是用来解决线性方程求解的工具
- 矩阵是高等代数中的常见工具，也常见于统计分析等应用数学学科中；矩阵在物理学和计算机科学中都有应用
- 矩阵的运算是数值分析领域的重要问题

# 矩阵



# 矩阵

| | | | | | |
|---|---|---|---|---|---|
| 1 | 2 | 3 | 4 | 5 | 6 |
| 7 | 8 | 9 | 10 | 11 | 12 |
| 13 | 14 | 15 | 16 | 17 | 18 |

$$\begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 7 & 8 & 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 & 17 & 18 \end{bmatrix}$$

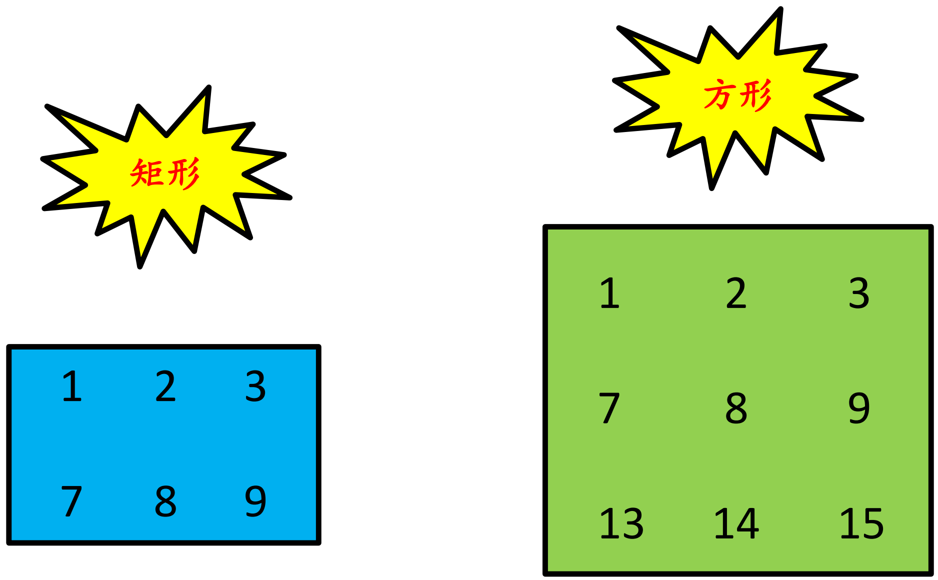
$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \cdots & \cdots & & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}$$

# 矩阵的定义

$$\begin{array}{ccc}
 & \begin{array}{cccc}\text{第 1 列} & \phantom{a_{12}} & \phantom{\cdots} & \text{第 n 列}\end{array} & \\
\mathbf{A} = & \left[\begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ a_{31} & a_{32} & \cdots & a_{3n} \\ \cdots & \cdots & & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{array}\right] & \begin{array}{l}\text{第 1 行} \\ \\ \\ \\ \text{第 m 行}\end{array}
\end{array}$$

- 由 m × n 个数 $a_{ij}$ (i = 1,2,...,m;  j = 1,2,...,n) 排成的 m 行 n 列的数表 A 就称为 m 行 n 列的**矩阵**
- 这 m × n 个数称作矩阵 A 的**元素**，元素 $a_{ij}$ 位于矩阵 A 的第 i 行第 j 列
- m × n 矩阵 A 可以记作 $A_{m×n}$，其中 m是行数，n是列数，m, n > 0

# 特殊矩阵




2行3列　　　3行3列

- 对于$A_{m\times n}$，如果 m = n，即矩阵的**行数与列数相等**，那么称A为**方阵**

# 矩阵中的概念

$A_{2\times3}$ → 2行3列矩阵

$A_{6\times6}$ → 6行6列矩阵 → 6阶方阵

$A_{6\times1}$ → 6行1列矩阵 → 6维列向量

$A_{1\times9}$ → 1行9列矩阵 → 9维行向量

- 行数与列数都等于 n 的矩阵称为 n 阶矩阵，又称做 **n 阶方阵**，可以记作 $A_n$
- 只有一行的矩阵 $A_{1\times n}$ 称为行矩阵，又叫**行向量**
- 同样，只有一列的矩阵 $A_{n\times1}$ 称为列矩阵，又叫**列向量**

# 矩阵中的概念

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \\ 13 & 14 & 15 & 16 \end{bmatrix}$$

主对角线

- 对于方阵，从左上角到右下角的直线，叫做**主对角线**，主对角线上的元素称为**主对角线元素**

# 特殊矩阵

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

矩阵的元素全部为0，称为**零矩阵**，用 O 表示

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

对于方阵，如果只有对角线元素为1，其余元素都为0，那么称为**单位矩阵**，一般用 I 或者 E 表示

$$\begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix}$$

对于方阵，不在对角线上的元素都为0，称为**对角矩阵**

# 矩阵的加法

- 把矩阵的对应位元素相加
- 矩阵的形状必须一致，即必须是**同型矩阵**

# 矩阵的加法

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} + \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 1+1 & 2+1 \\ 3+1 & 4+1 \end{bmatrix} = \begin{bmatrix} 2 & 3 \\ 4 & 5 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} + \begin{bmatrix} 9 & 8 & 7 \\ 6 & 5 & 4 \\ 3 & 2 & 1 \end{bmatrix} = \begin{bmatrix} 1+9 & 2+8 & 3+7 \\ 4+6 & 5+5 & 6+4 \\ 7+3 & 8+2 & 9+1 \end{bmatrix} = \begin{bmatrix} 10 & 10 & 10 \\ 10 & 10 & 10 \\ 10 & 10 & 10 \end{bmatrix}$$

# 矩阵的乘法

1. 数与矩阵相乘

- 数值与矩阵每一个元素相乘

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \times 2 = \begin{bmatrix} 1 \times 2 & 2 \times 2 \\ 3 \times 2 & 4 \times 2 \end{bmatrix} = \begin{bmatrix} 2 & 4 \\ 6 & 8 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \times 3 = \begin{bmatrix} 1 \times 3 & 2 \times 3 & 3 \times 3 \\ 4 \times 3 & 5 \times 3 & 6 \times 3 \\ 7 \times 3 & 8 \times 3 & 9 \times 3 \end{bmatrix} = \begin{bmatrix} 3 & 6 & 9 \\ 12 & 15 & 18 \\ 21 & 24 & 27 \end{bmatrix}$$

# 矩阵的乘法

2. 矩阵与矩阵相乘

- 左矩阵的每一行与右矩阵的每一列，对应每一个元素相乘

保留并喜结连理

$$A_{\text{行}\times\text{列}} \times B_{\text{行}\times\text{列}} = C_{\text{行}\times\text{列}}$$

相等并同归于尽

# 矩阵的乘法

## 2. 矩阵与矩阵相乘

- $A \times B$，那么有 A 矩阵 $m \times n$，B 矩阵 $n \times k$，要求左侧矩阵的列数 n，必须等于右侧矩阵的行数 n，结果矩阵 C 为 $m \times k$ 矩阵。

| | | | | |
|---|---|---|---|---|
| $A_{2\times3}$ | $B_{2\times3}$ | $A_{2\times3}$ | $B_{2\times3}$ | ✗ |
| $A_{2\times3}$ | $B_{3\times2}$ | $A_{2\times3}$ | $B_{3\times2}$ | $C_{22}$ |
| $A_{3\times3}$ | $B_{3\times3}$ | $A_{3\times3}$ | $B_{3\times3}$ | $C_{33}$ |
| $A_{2\times2}$ | $B_{2\times2}$ | $A_{2\times2}$ | $B_{2\times2}$ | $C_{22}$ |

$$\begin{bmatrix} \boxed{1 \quad 2 \quad 3} \\ 4 \quad 5 \quad 6 \\ 7 \quad 8 \quad 9 \end{bmatrix} \times \begin{bmatrix} \boxed{\begin{matrix}1\\2\\3\end{matrix}} & \begin{matrix}4\\5\\6\end{matrix} & \begin{matrix}7\\8\\9\end{matrix} \end{bmatrix} = \begin{bmatrix} \boxed{1 \times 1 + 2 \times 2 + 3 \times 3 = 14} & & \\ & & \\ & & \end{bmatrix}$$

$$\begin{bmatrix} \boxed{1 \quad 2 \quad 3} \\ 4 \quad 5 \quad 6 \\ 7 \quad 8 \quad 9 \end{bmatrix} \times \begin{bmatrix} \begin{matrix}1\\2\\3\end{matrix} & \boxed{\begin{matrix}4\\5\\6\end{matrix}} & \begin{matrix}7\\8\\9\end{matrix} \end{bmatrix} = \begin{bmatrix} 14 & \boxed{1 \times 4 + 2 \times 5 + 3 \times 6 = 32} & \\ & & \\ & & \end{bmatrix}$$

$$\begin{bmatrix} \boxed{1 \quad 2 \quad 3} \\ 4 \quad 5 \quad 6 \\ 7 \quad 8 \quad 9 \end{bmatrix} \times \begin{bmatrix} \begin{matrix}1\\2\\3\end{matrix} & \begin{matrix}4\\5\\6\end{matrix} & \boxed{\begin{matrix}7\\8\\9\end{matrix}} \end{bmatrix} = \begin{bmatrix} 14 & 32 & \boxed{1 \times 7 + 2 \times 8 + 3 \times 9 = 50} \\ & & \\ & & \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} =$$

| 14 | 32 | 50 |
|---|---|---|
| 4 × 1 + 5 × 2 + 6 × 3 = 32 | | |
| | | |

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} =$$

| 14 | 32 | 50 |
|---|---|---|
| 32 | 4 × 4 + 5 × 5 + 6 × 6 =77 | |
| | | |

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} =$$

| 14 | 32 | 50 |
|---|---|---|
| 32 | 77 | 4 × 7 + 5 × 8 + 6 × 9 = 122 |
| | | |

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ \boxed{7 \quad 8 \quad 9} \end{bmatrix} \times \begin{bmatrix} \boxed{1} & 4 & 7 \\ \boxed{2} & 5 & 8 \\ \boxed{3} & 6 & 9 \end{bmatrix} =$$

| 14 | 32 | 50 |
|---|---|---|
| 32 | 77 | 122 |
| 7 × 1 + 8 × 2 + 9 × 3 = 50 | | |

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ \boxed{7 \quad 8 \quad 9} \end{bmatrix} \times \begin{bmatrix} 1 & \boxed{4} & 7 \\ 2 & \boxed{5} & 8 \\ 3 & \boxed{6} & 9 \end{bmatrix} =$$

| 14 | 32 | 50 |
|---|---|---|
| 32 | 77 | 122 |
| 50 | 7 × 4 + 8 × 5 + 9 × 6 = 122 | |

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ \boxed{7 \quad 8 \quad 9} \end{bmatrix} \times \begin{bmatrix} 1 & 4 & \boxed{7} \\ 2 & 5 & \boxed{8} \\ 3 & 6 & \boxed{9} \end{bmatrix} =$$

| 14 | 32 | 50 |
|---|---|---|
| 32 | 77 | 122 |
| 50 | 122 | 7 × 7 + 8 × 8 + 9 × 9 = 194 |

# 矩阵的乘法

$$\begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \times \begin{bmatrix} 1 & 4 & 7 \\ 2 & 5 & 8 \\ 3 & 6 & 9 \end{bmatrix} =$$

| 14 | 32 | 50 |
|---|---|---|
| 32 | 77 | 122 |
| 50 | 122 | 194 |

- 规则：一行乘一列，行定列移动，列尽下一行

$$\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{pmatrix} \begin{pmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \\ b_{31} & b_{32} \end{pmatrix}$$

$$= \begin{pmatrix} a_{11}b_{11} + a_{12}b_{21} + a_{13}b_{31} & a_{11}b_{12} + a_{12}b_{22} + a_{13}b_{32} \\ a_{21}b_{11} + a_{22}b_{21} + a_{23}b_{31} & a_{21}b_{12} + a_{22}b_{22} + a_{23}b_{32} \end{pmatrix}$$

# 矩阵的乘法

$$\begin{bmatrix}1 & 2\\3 & 4\end{bmatrix} \times \begin{bmatrix}2 & 2\\2 & 2\end{bmatrix} = \begin{bmatrix}1\times2+2\times2 & 1\times2+2\times2\\3\times2+4\times2 & 3\times2+4\times2\end{bmatrix} = \begin{bmatrix}6 & 6\\14 & 14\end{bmatrix}$$

$$\begin{bmatrix}1 & 2 & 3\\4 & 5 & 6\\7 & 8 & 9\end{bmatrix} \times \begin{bmatrix}1 & 2 & 3\\3 & 2 & 1\\1 & 2 & 3\end{bmatrix}$$

$$= \begin{bmatrix}1\times1+2\times3+3\times1 & 1\times2+2\times2+3\times2 & 1\times3+2\times1+3\times3\\4\times1+5\times3+6\times1 & 4\times2+5\times2+6\times2 & 4\times3+5\times1+6\times3\\7\times1+8\times3+9\times1 & 7\times2+8\times2+9\times2 & 7\times3+8\times1+9\times3\end{bmatrix}$$

$$= \begin{bmatrix}10 & 12 & 14\\25 & 30 & 35\\40 & 48 & 56\end{bmatrix}$$

# 矩阵的转置

$$\begin{bmatrix} 1 & 2 & 3 & 4 \\ 5 & 6 & 7 & 8 \\ 9 & 10 & 11 & 12 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 5 & 9 \\ 2 & 6 & 10 \\ 3 & 7 & 11 \\ 4 & 8 & 12 \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \rightarrow \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \cdots & \cdots & & \cdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}$$

- 把矩阵 A 的行换成相同序数的列，得到一个新矩阵，叫做 A 的**转置矩阵**，记作 $A^T$
- 行变列，列变行
- A 为 m × n 矩阵，转置之后为 n × m 矩阵

# 矩阵的运算法则

## 加法

- $A + B = B + A$
- $(A + B) + C = A + (B + C)$

## 减法

- $A - B = A + B \times (-1)$
- $A - A = A + (-A) = O$

## 乘法

- $(\lambda\mu) A = \lambda (\mu A)$
- $(\lambda + \mu) A = \lambda A + \mu A$
- $\lambda (A + B) = \lambda A + \lambda B$
- $(AB) C = A (BC)$
- $\lambda (AB) = (\lambda A) B = A (\lambda B)$
- $A (B + C) = AB + AC$

  $(B + C) A = BA + CA$

## 转置

- $(A^T)^T = A$
- $(A + B)^T = A^T + B^T$
- $(\lambda A)^T = \lambda A^T$
- $(AB)^T = B^T A^T$

# 矩阵的逆

- 对于 n 阶方阵 A，如果有一个 n 阶方阵 B，使得

$$AB = BA = E$$

就称矩阵 A 是**可逆的**，并把 B 称为 A 的**逆矩阵**

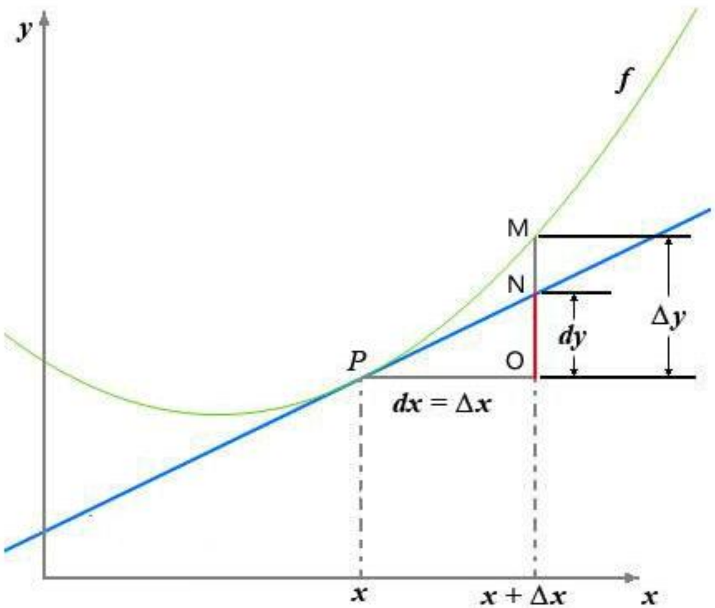
- A 的逆矩阵记作 $A^{-1}$，如果 $AB = BA = E$，则 $B = A^{-1}$

$$\begin{bmatrix} 1 & 2 & 3 \\ 2 & 2 & 1 \\ 3 & 4 & 3 \end{bmatrix} \times \begin{bmatrix} 1 & 3 & -2 \\ -3/2 & -3 & 5/2 \\ 1 & 1 & -1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

# 微积分基本知识

- 什么是导数
- 偏导数
- 方向导数和梯度
- 凸函数和凹函数

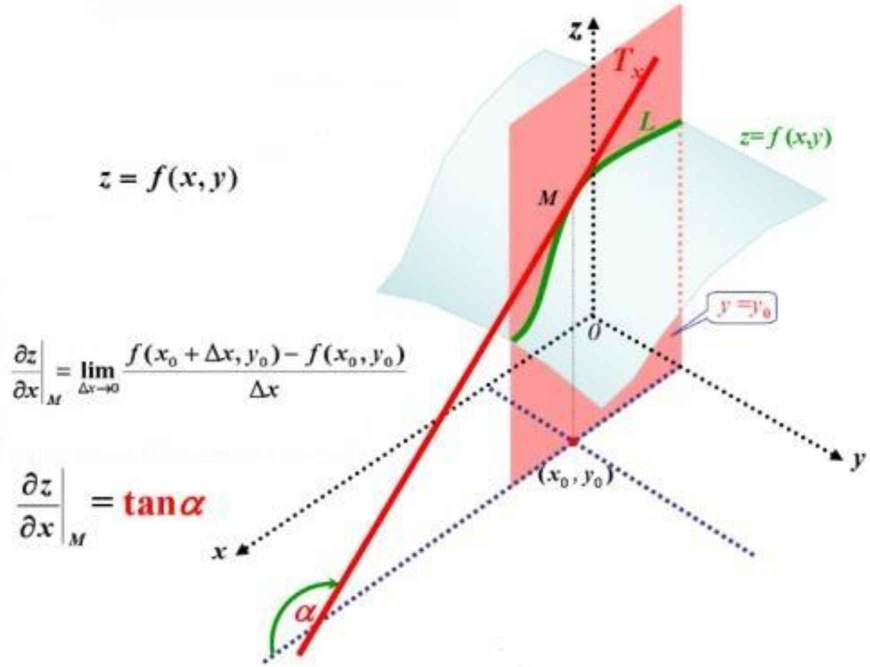
# 导数




$$f'(x_0) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

- 导数反映的是函数 y = f(x) 在某一点处沿 x 轴正方向的变化率
- 在x轴上某一点处，如果 f’(x)>0，说明f(x)的函数值在x点沿x轴正方向是趋于增加的；如果 f’(x)<0，说明f(x)的函数值在x点沿x轴正方向是趋于减少的

# 偏导数

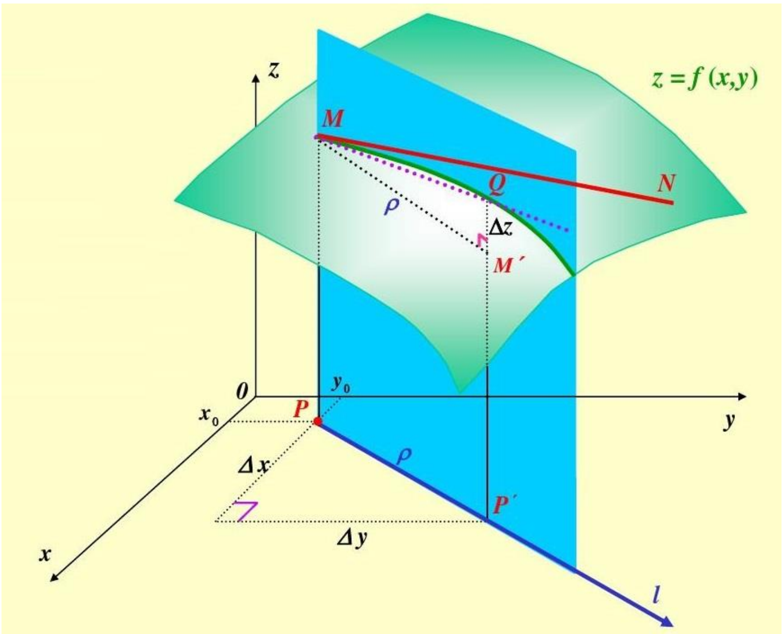



- 导数与偏导数本质是一致的，都是当自变量的变化量趋于0时，函数值的变化量与自变量变化量比值的极限
- 偏导数也就是函数在某一点上沿某个坐标轴正方向的的变化率

$$\frac{\partial}{\partial x_j} f(x_0, x_1, \dots, x_n) = \lim_{\Delta x \to 0} \frac{\Delta y}{\Delta x} = \lim_{\Delta x \to 0} \frac{f(x_0, \dots, x_j + \Delta x, \dots, x_n) - f(x_0, \dots, x_j, \dots, x_n)}{\Delta x}$$

- 导数指的是一元函数中，函数 y=f(x) 在某一点处沿 x 轴正方向的变化率；
  而偏导数，指的是多元函数中，函数 $y=f(x_1,x_2,\cdots,x_n)$ 在某一点处沿某一坐标轴（$x_1,x_2,\cdots,x_n$）正方向的变化率

# 方向导数



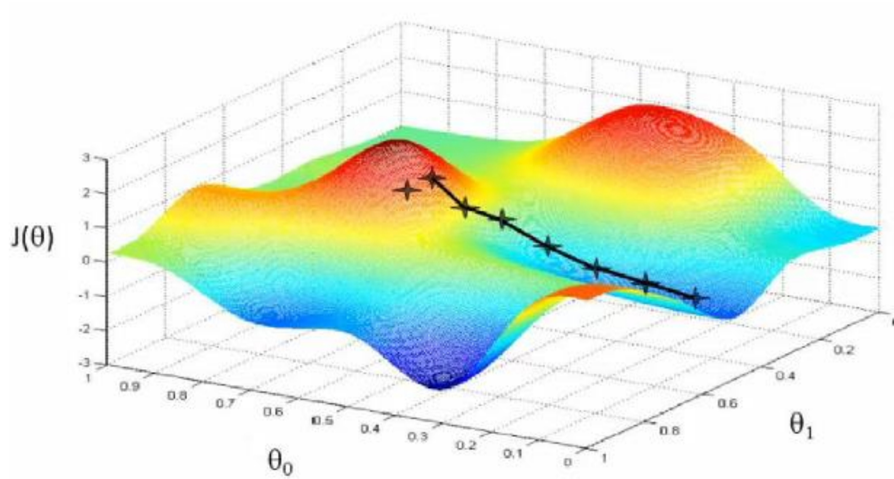

- 函数某一点在某一趋近方向（向量方向）上的导数值
- 方向导数就是函数在除坐标轴正方向外，其他特定方向上的变化率

$$\frac{\partial}{\partial l} f(x_0, x_1, \dots, x_n) = \lim_{\rho \to 0} \frac{\Delta y}{\Delta x} = \lim_{\rho \to 0} \frac{f\left(x_0 + \Delta x_0, \dots, x_j + \Delta x_j, \dots, x_n + \Delta x_n\right) - f\left(x_0, \dots, x_j, \dots, x_n\right)}{\rho}$$

$$\rho = \sqrt{(\Delta x_0)^2 + \cdots + \left(\Delta x_j\right)^2 + \cdots + (\Delta x_n)^2}$$

# 梯度（Gradient）

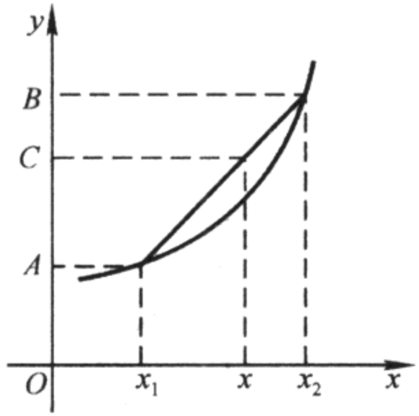
问题：函数在变量空间的某一点处，沿着哪一个方向有最大的变化率？



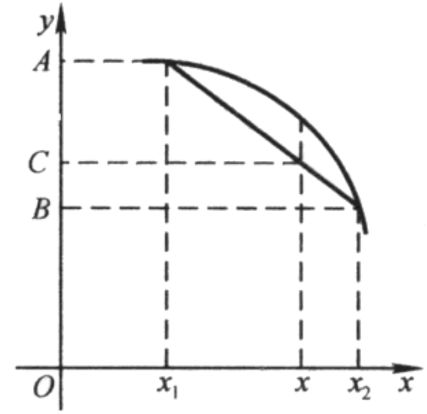

$$gradf(x_0, x_1, \ldots, x_n) = \left(\frac{\partial f}{\partial x_0}, \ldots, \frac{\partial f}{\partial x_j}, \ldots, \frac{\partial f}{\partial x_n}\right)$$

- 定义：函数在某一点的梯度是这样一个向量，它的方向与取得最大方向导数的方向一致，而它的模为方向导数的最大值
- 梯度是一个向量，即有方向、有大小；
- 梯度的方向是最大方向导数的方向；梯度的值是最大方向导数的值

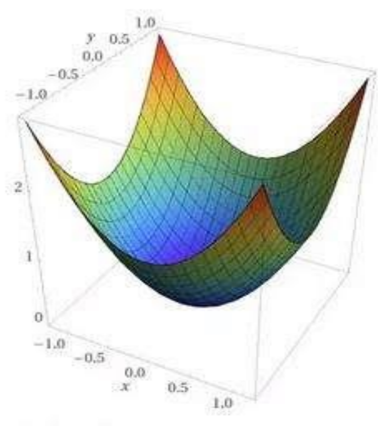
# 凸函数和凹函数

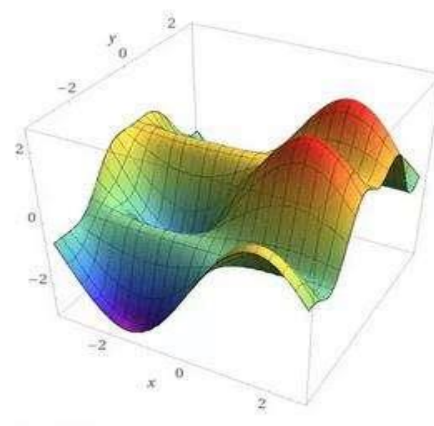



凸函数




凹函数





# 概率统计基础知识

- 常用统计变量
- 常见概率分布
- 重要概率公式

# 常用统计变量

- 样本均值

$$E(X) = \overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$$

- 样本方差

$$D(X) = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i^2 - n\overline{X})$$

- 样本标准差

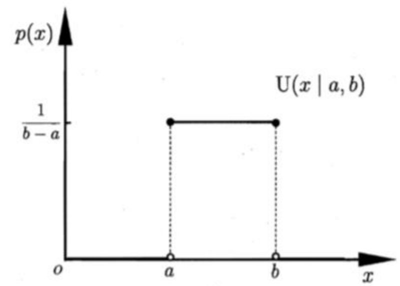
$$\sqrt{D(X)} = S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})^2}$$

# 常见概率分布

- 均匀分布

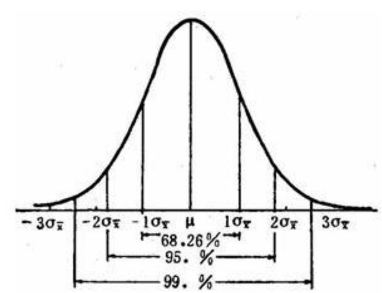
$$f(x) = \frac{1}{b-a}, a < x < b$$

$$f(x) = 0, else$$



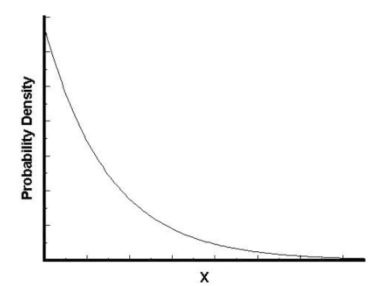

- 正态分布（高斯分布）

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$




- 指数分布

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0, \\ 0, & x < 0, \end{cases}$$



# 重要概率公式

- 条件概率公式

$$P(B|A) = \frac{P(AB)}{P(A)}$$

- 全概率公式

$$P(A) = P(A|B_1)P(B_1) + P(A|B_2)P(B_2) + \ldots + P(A|B_n)P(B_n)$$

- 贝叶斯公式

$$P(B_i|A) = \frac{P(A|B_i)P(B_i)}{\sum_{j=1}^{n} P(A|B_j)P(B_j)}$$

# Q & A