第二章数学知识

2.1 微积分和线性代数

2.1.1 导数

- 导数大于0 时函数单调增，导数小于0 时函数单调减，在极值点处导数必定为0。
- 二阶导数决定函数的凹凸性。如果二阶导数大于0 ， 则函数为凸函数；如果二阶导数小于0 ， 则为凹函数
- 导数等于0 的点称为函数的驻点
- 二阶导数等于0 的点称为函数的拐点

一元函数的极值判别法：

- 在驻点处，
  - 如果二阶导数大于0 ， 则为函数的极小值点；
- 如果二阶导数小于0 ， 则为极大值点。
- 如果二阶导数等于0，则情况不定。

2.1.2 向量与矩阵

- 向量的L- P 范数是一个标量，定义
- 矩阵乘法
  - **(AB)C = A(BC)**
  - **A(B + C) =AB +AC**
  - **（A+ B ）C= AC+BC**
  - **AB$\neq$ BA**
- n阶矩阵A，n阶矩阵B
  - $AA^{-1} = I$
  - $A^{-1}A = I$
  - $(AB)^{-1} = B^{-1}A^{-1}$
  - $(A^{-1})^{-1} = A$
  - $(A^T)^{-1} = (A^{-1})^T$

2.1.3 偏导数与梯度

- 多元函数$f(x_1, \ldots, x_n)$,对自变量x_i的偏导数定义为
  $\frac{\partial f}{\partial x_i} = lim_{\Delta x_i \to 0} \frac{f(x_1,\ldots,x_i+\Delta x_i,x_n)-f(x_1,\ldots,x_i,\ldots,x_n)}{\Delta x_i}$
- 多元函数的梯度定义为,多元函数对各个自变量偏导数形成的向量

  梯度是导数对多元函数的推广

  $\nabla f(x) = (\frac{\partial f}{\partial x_1}, \ldots, \frac{\partial f}{\partial x_n})^T$
- 梯度和函数的单调性、极值有关
  - 可导函数在某一点处取得极值的必要条件是梯度为0
  - 梯度为0 的点称为函数的驻点
- 一般情况下，混合二阶偏导数与求导次序无关

$$\frac{\partial^2 f}{\partial x \partial y}=\frac{\partial^2 f}{\partial y \partial x}$$

2.1.4 雅克比矩阵

- 雅克比矩阵定义为输出向量的每个分量对输入向量的每个分量的偏导数构成的矩阵：
- 例子

  $u=x^2+2xy+z$

  $v=x-y^2+z^2$

  $$\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} & \frac{\partial u}{\partial z} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} & \frac{\partial v}{\partial z} \end{bmatrix}=\begin{bmatrix} 2x+2y & 2x & 1 \\ 1 & -2y & 2z \end{bmatrix}$$

2.1.5 Hessian 矩阵

- Hessian 矩阵是一个对称矩阵

  它可以看作二阶导数对多元函数的推广
- 多元函数极值判别法:
  - 多元函数在点M 的梯度为0， 即M 是函数的驻点
    1. 如果Hes s i a n 矩阵正定，函数在该点有极小值。
    2. 如果H ess i a n 矩阵负定，函数在该点有极大值
    3. 如果Hessian 矩阵不定，则不是极值点
- 判定矩阵正定的常用方法
  - 矩阵的特征值全大于0
  - 矩阵的所有顺序主子式都大于0
  - 矩阵合同于单位阵**I**
- Hessian 矩阵正定性与多元函数的凹凸性有关
  - 如果Hessian矩阵半正定， 则函数为凸函数；
- 如果Hessian 矩阵正定，则为严格凸函数。

2.1.6 泰勒展开

- 多元函数的泰勒展开公式为

  $f(x)=f(x_0)+(\nabla f(x_0))^T(x-x_0)+\frac{1}{2}(x-x_0)+o(\|x-x_0\|^2)$

2.1.7 行列式

- 如果矩阵A和B是尺寸相同的n阶矩阵，则有$|AB|=|A||B|$
- 如果矩阵可逆，$|A^{-1}|=|A|^{-1}$
- 矩阵与标量乘法，$|aA|=a^n|A|$
- $|A^T|=|A|$

2.1.8 特征值与特征向量

- 矩阵的迹定义为主对角线元素之和：$tr(A)=\sum_{i=1}^n a_{ii}$
- 矩阵所有特征值的和为矩阵的迹：$\sum_{i=1}^n \lambda_i=tr(A)$
- 矩阵所有特征值的积为矩阵的行列式：$\prod_{i=1}^n \lambda_i=|A|$

1. 正交矩阵

- 定义：一个n阶矩阵满足：$P^{-1}=P^T$则称为正交矩阵

- 定义2：行向量之间相互正交，相同行的内积为1，不同行内积为0
- 定义3：列向量之间相互正交，相同列的内积为1，不同列内积为0
- 性质：
    - 正交矩阵的行列式为1或-1

2. 矩阵对角化（A为方阵，有条件）

- 定义：对于一个n阶矩阵A的对角化，$P^{-1}AP=\Lambda$
    - 矩阵P为正交矩阵，称为对角化旋转矩阵，P的列为矩阵A的正交化特征向量
    - 矩阵Λ为对角矩阵，的对角元素为矩阵A的特征值
- 矩阵A可以对角化的充分必要条件是存在n个线性无关的特征向量
- 实对称矩阵一定可以对角化

2.1.9 奇异值分解（SVD)(不一定是方阵,无条件）

- 定义：A是一个m*n的矩阵，存在$A=U\Sigma V^T$
    - U为m*m的正交矩阵，其列称为矩阵A的左奇异向量。

      U的列为AA^T的特征向量
    - Σ为m*n的对角矩阵，除了主对角线以外，其他元素都是0.

      主对角线为AA^T或A^TA的特征值$\sqrt{\lambda_{ii}}$从大到小排
    - V为n*n的正交矩阵，其行称为矩阵A的右奇异向量

      U的列为A^TA的特征向量

2.1.10 二次型

2.1.11 向量与矩阵求导

- 多元一次函数$f(\mathbf{x})=\mathbf{w}^T\mathbf{x}$　　　(只有$X_1,...,X_n$)

  梯度: $\nabla\mathbf{w}^T\mathbf{x}=\mathbf{w}$
- 多元二次函数：$f(\mathbf{x})=\mathbf{x}^T\mathbf{A}\mathbf{x}$　　　(只有X_1^2,X_1X_2之类的二次项)

  梯度：　$\nabla\mathbf{x}^T\mathbf{A}\mathbf{x}=(\mathbf{A}+\mathbf{A}^T)\mathbf{x}$

  二阶梯度：$\nabla^2\mathbf{x}^T\mathbf{A}\mathbf{x}=\mathbf{A}+\mathbf{A}^T$

  如果A是对称矩阵：

  梯度：　$\nabla\mathbf{x}^T\mathbf{A}\mathbf{x}=2\mathbf{A}\mathbf{x}$

  二阶梯度：$\nabla^2\mathbf{x}^T\mathbf{A}\mathbf{x}=2\mathbf{A}$