

内 容 简 介

本书记录了作者在山大数院的三年所学的数学知识，简明扼要地列举出每个学科所必须掌握的定理等重要知识点。此书不宜作为初学某一科目的参考资料，而适合已学完部分内容者查漏补缺。

根据山大本科开课顺序以及个人自学进度，大致收录下列学科：

大一上：数学分析 1，高等代数 1，解析几何

大一下：数学分析 2，高等代数 2

大二上：数学分析 3，复变函数，常微分方程

大二下：实变函数，偏微分方程，概率论

大三上：机器学习、数字图像处理、数理统计、最优化方法

大三下：时间序列分析、数据库系统、数据结构

目 录

第一章 数学分析	1
1.1 预备知识	2
1.1.1 集合	2
1.1.2 映射	3
1.2 极限	5
1.2.1 实数系连续性	5
1.2.2 数列极限	6
1.2.3 函数极限	8
1.3 连续函数	9
1.3.1 点态连续	9
1.3.2 闭区间上的连续函数	10
1.3.3 一致连续	11
1.4 一元函数微分学	11
1.4.1 微分和导数的概念	12
1.4.2 求导法则	13
1.4.3 微分中值定理	14
1.4.4 泰勒展开	15

1.5	一元函数积分学	16
1.5.1	不定积分	16
1.5.2	定积分	17
1.5.3	广义积分	19
1.6	无穷级数	21
1.6.1	数项级数	22
1.6.2	函数项级数	23
1.6.3	傅里叶级数	25
1.7	多元函数微分学	27
1.7.1	欧几里得空间	27
1.7.2	多元连续函数	29
1.7.3	偏导数与全微分	30
1.7.4	多元函数求导法则	32
1.7.5	多元函数微分学的应用	33
1.8	多元函数积分学	35
1.8.1	重积分	36
1.8.2	曲线积分	39
1.8.3	曲面积分	40
1.8.4	流形上的 Stokes 公式	41
1.8.5	场论	44
1.8.6	含参变量积分	45
第二章	高等代数	47
2.1	多项式	48
2.1.1	数域与一元多项式	48

2.1.2	一元多项式的因式	48
2.1.3	多项式的因式分解	49
2.1.4	有理系数多项式	49
2.2	向量与线性方程组	50
2.2.1	线性方程组	50
2.2.2	向量空间	51
2.2.3	线性方程组解的结构	52
2.3	矩阵与行列式	53
2.3.1	矩阵的概念	53
2.3.2	行列式	54
2.3.3	在解线性方程组中的应用	56
2.3.4	矩阵的性质	56
2.4	二次型	57
2.4.1	二次型的矩阵表示	58
2.4.2	二次型的化简	58
2.5	线性空间	59
2.5.1	线性空间的定义	59
2.5.2	基与坐标	60
2.5.3	基变换与坐标变换	60
2.5.4	线性子空间	61
2.6	线性变换	62
2.6.1	映射与变换	62
2.6.2	线性变换的运算	63
2.6.3	线性变换的矩阵表示	63
2.6.4	特征值与特征向量	64

2.6.5	线性变换的值域与核	65
2.6.6	不变子空间	66
2.7	λ -矩阵	66
2.7.1	基本概念	66
2.7.2	标准形	66
2.7.3	相似不变量	67
2.7.4	数字矩阵的相似标准形	68
2.8	欧几里得空间	69
2.8.1	欧氏空间与内积	69
2.8.2	标准正交基	71
2.8.3	正交变换	71
2.8.4	实对称矩阵的正交对角化	72
第三章	概率论与数理统计	73
3.1	随机事件与概率	74
3.1.1	随机事件	74
3.1.2	概率	74
3.1.3	概率的性质	75
3.1.4	条件概率	75
3.2	随机变量及其分布	76
3.2.1	随机变量	76
3.2.2	常用分布及概率密度函数	77
3.2.3	数字特征	77
3.2.4	随机变量函数的分布	78
3.3	多元随机变量及其分布	78

3.3.1	多元随机变量	78
3.3.2	边缘分布	79
3.3.3	多元随机变量函数的分布	79
3.3.4	多元随机变量的特征数	80
3.3.5	条件分布与条件期望	80
3.4	大数定律和中心极限定理	81
3.4.1	随机变量的特征函数	81
3.4.2	大数定律	82
3.4.3	中心极限定理	82
3.5	数理统计基本概念	83
3.5.1	统计学基本思想	83
3.5.2	常用统计量	84
3.5.3	抽样分布	84
3.5.4	利用抽样分布统计推断	85
3.5.5	充分统计量	86
3.6	参数估计	86
3.6.1	矩估计	86
3.6.2	最大似然估计	87
3.6.3	点估计的评价标准	88
3.6.4	贝叶斯估计	88
3.6.5	区间估计	89
3.7	假设检验	91
3.7.1	基本思想	91
3.7.2	正态总体假设检验	92
3.7.3	广义似然比检验	93

3.7.4	拟合优度检验	94
3.7.5	正态性检验	94
3.7.6	游程检验	94
3.8	方差分析	95
3.8.1	基本思想	95
3.8.2	单因素方差分析	96
3.8.3	方差齐性检验	96
3.9	回归分析	97
3.9.1	基本思想	97
3.9.2	回归系数的最小二乘估计	97
3.9.3	区间估计与预测	98
3.9.4	显著性检验	98
3.9.5	多元线性回归	99
3.9.6	非线性回归	99

第一章 数学分析

Mathematical Analysis

数学分析是大一新生所修的重要学科基础课, 相比非数学专业更强调证明, 对收敛性的讨论篇幅较大, 与大二的实变函数课程联系紧密. 数分是今后多门专业课的先修课程: 积分学应用于概率论对随机变量的研究; 对积分的进一步研究 (Lebesgue 积分) 是实变函数的重要内容; Fourier 变换和多元函数积分学是偏微分方程必不可少的工具... 山大主选教材为陈纪修的《数学分析》(第三版), 在此基础上结合卓里奇的数学分析教程, 对共计三个学期的数分课程进行完整的内容回顾.

数分 1 的重点: 极限与连续概念, 一元函数微分学, 微分中值定理

数分 2 的重点: 一元函数积分学, 数项级数和函数项级数, 广义积分

数分 3 的重点: 多元函数微分学, 含参变量积分, 多元函数积分学 (重积分, 曲线与曲面积分)

1.1 预备知识

这部分主要涉及集合与函数概念, 内容不多且不难, 可以认为是从高中数学到数分的过渡, 更可以认为是数学各个分支的基石.

集合论是高等数学的核心, 由此衍生出基 (tu) 础 (tou) 数学和计算机科学的区别: 一个研究连续集合, 比如实数域, 复数域等具有连续势集合上的映射, 另一个更偏向离散集合, 也就是有穷集和可列集上的映射. 从前者开始诞生实分析, 复分析, 傅里叶分析, 泛函分析等各大分析, 后者则衍生出图论, 组合数学, 数据结构等计算机科学分支. 认清这一点后, 我们便可以用一句话概括数学分析干了啥: 研究实数域或 n 维欧氏空间到实数域上的映射. 同时, 集合论又是各大学科的基础 (笔者在数分, 实变, 离散数学三门课上过三遍集合论...), 故不可轻敌.

映射就是数学分析的研究主体. 注意到我们只研究欧氏空间到实数域的映射, 也就是实变量函数, 我们可以归纳出这一类函数的表示方法和基本性质, 同时温习一下高中数学内容.

1.1.1 集合

1. 定义: 具有某种特定性质的对象总体。
2. 集合与元素间的关系: 若元素 a 在集合 A 内, 则称 a 属于 A , 记作 $a \in A$.
3. 集合之间的关系: 若集合 A 的所有元素同时为集合 B 内的元素, 则称 A 包含于 B , 记作 $A \subset B$; 若 $A \subset B$, 且 $B \subset A$, 则称集合 A 与集合 B 相等, 记作 $A = B$.
4. 集合的运算:
 - (1) 交: $a \in A \cap B$ 当且仅当 $a \in A$ 且 $a \in B$, 称 $A \cap B$ 为集合 A 和集合 B

的交；

(2) 并： $a \in A \cup B$ 当且仅当 $a \in A$ 或 $a \in B$ ，称 $A \cup B$ 为集合 A 和集合 B 的并；

(3) 补： $a \in A^C$ 当且仅当 $a \notin A$ ，称 A^C 为集合 A 的补；

(4) 差： $a \in A - B$ 当且仅当 $a \in A$ 且 $a \notin B$ ，称 $A - B$ 为集合 A 和集合 B 的差。

5. 集合运算的性质：

(1) 结合律： $A \cap (B \cap C) = (A \cap B) \cap C$ ； $A \cup (B \cup C) = (A \cup B) \cup C$ ；

(2) 交换律： $A \cap B = B \cap A$ ； $A \cup B = B \cup A$ ；

(3) 分配律： $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$ ； $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ ；

(4) 德摩根律： $(A \cap B)^C = A^C \cup B^C$ ； $(A \cup B)^C = A^C \cap B^C$ 。

6. 集合的势：当集合有有限个元素时，集合的元素个数称为集合的势；当集合有无限个元素，但可以按照某种规律排成一系列时，称集合具有可列势。

7. 笛卡尔积：集合 $A \times B = \{(x, y) | x \in A, y \in B\}$ 称为集合 A 和集合 B 的笛卡尔积。

1.1.2 映射

1. 定义：若存在集合 X 与集合 Y 间的对应关系 f ，使得任意 $x \in X$ ，存在唯一的 $y \in Y$ 使得 $f(x) = y$ ，则称对应关系 f 为集合 X 到集合 Y 上的映射，记作 $f: X \rightarrow Y$ 。此时称 x 为 y 的原像， y 为 x 的像。

2. 特殊的映射：

(1) 单射：若 $f(x) = f(y)$ 当且仅当 $x = y$ ，即一个像唯一对应一个原像，则称 f 为单射；

(2) 满射：若对任意 $y \in Y$ ，均存在 $x \in X$ ，使得 $f(x) = y$ ，则称 f 为满射；

(3) 双射: 若 f 既为单射又为满射, 则称 f 为双射, 又称为一一对应。

3. 函数: 若映射 f 的定义域和值域均限制为数域, 则称 f 为函数, 此时 x 称为自变量, y 称为因变量; 特殊地, 当 $X \subset R, Y = R$ 时, 称 f 为一元实函数。

4. 基本初等函数:

(1) 常数函数: $y = c$, 其中 c 为常数;

(2) 幂函数: $y = x^a$;

(3) 指数函数: $y = a^x$, 其中 $a > 0$ 且 $a \neq 1$;

(4) 对数函数: $y = \log_a x$, 其中 $a > 0$ 且 $a \neq 1$;

(5) 三角函数: $y = \sin x, y = \cos x, y = \tan x, y = \cot x$;

(6) 反三角函数: $y = \arcsin x, y = \arccos x, y = \arctan x$ 。

5. 参数方程: 形如

$$\begin{cases} x = x(t) \\ y = y(t) \end{cases}, t \in [a, b]$$

用参数 t 间接表示自变量和因变量关系的方程组称为参数方程。

6. 隐函数: 形如 $F(x, y) = 0$ 的函数称为隐函数, 通常解不出 $y = f(x)$ 的显式函数表达式。

7. 函数的特殊性质:

(1) 有界: 若存在 $M > 0$, 使得任意 $x \in X$, 均有 $|f(x)| \leq M$, 则称 $f(x)$ 为有界函数。

(2) 单调: 若对任意满足 $x_1 < x_2$ 的 $x_1, x_2 \in X$, 均有 $f(x_1) \leq f(x_2)$, 则称 $f(x)$ 在 X 上单调递增, 等号无法成立时称为严格单调递增; 若对任意满足 $x_1 < x_2$ 的 $x_1, x_2 \in X$, 均有 $f(x_1) \geq f(x_2)$, 则称 $f(x)$ 在 X 上单调递减, 等号无法成立时称为严格单调递减。

(3) 奇偶: 当 X 关于原点对称时, 若对任意 $x \in X$, 均有 $f(x) = f(-x)$, 则称 $f(x)$ 在 X 上为偶函数; 若对任意 $x \in X$, 均有 $f(x) = -f(-x)$, 则称 $f(x)$ 在

X 上为奇函数。

(4) 周期: 若存在某一正数 $T \in \mathbf{R}$, 使得任意 $x \in X$, 均有 $f(x+T) = f(x)$, 则称 $f(x)$ 为周期函数, T 为周期; 符合条件的最小正数 T 称为 $f(x)$ 的最小正周期。

1.2 极限

极限是数学分析最基本的概念, 研究无限逼近时的数列和函数性质。由于极限是涉及无穷的概念, 因此时常和直观感受不符。学习极限概念时, 最重要的就是掌握如何用最严谨的 $\varepsilon - N$ 语言描述, 而不是靠着想当然来回答问题。

为了研究实数系上的极限, 必须证明实数系的稠密性, 即无限逼近的过程中, 任意时刻的位置仍在实数轴上。在此基础上, 根据“姚多近优多近”的思维方式取定和 ε 相关的 N , 严谨地推导极限的相关性质。

在引入极限的严谨定义后, 围绕数列和函数的极限产生了实数系完备性定理, 分别是确界存在性定理、单调有界准则、柯西收敛准则、闭区间套定理和 Bolzano-Weierstrass 定理。这五条定理是实分析严谨理论的基石, 初学时不妨尝试着翻来覆去推导几遍。

温馨提示: 从极限开始, 会有很多需要自己动手证明的命题或者定理, 用上标 “*” 表示。请尽量尝试独立推导和证明。

1.2.1 实数系连续性

1. 数系的扩充:

(1) 整数系 \mathbb{N} : 所有整数构成的集合, 具有对加法运算和乘法运算的封闭性。

(2) 有理数系 \mathbb{Q} : 所有形如 $\frac{p}{q}$ 形式的有理数构成的集合, 其中 p, q 均为非零整数。

(3) 实数系 \mathbb{R} : 所有有理数和无理数构成的集合。

2. 实数系的连续性: $\forall x_1, x_2 \in \mathbb{R}, \exists y \in \mathbb{R}, x_1 < y < x_2$;

3. 最值: 对实数集 $A \in \mathbb{R}$, 若 $\exists x \in A, \forall y \in A, x \geq y$, 则称 x 为 A 的最大值; 若 $\exists x \in A, \forall y \in A, x \leq y$, 则称 x 为 A 的最小值, 记作 $x = \min A$ 。

4. 确界: 若集合 A 的所有上界组成的集合为 U , 则称 $\min U$ 为集合 A 的上确界, 记作 $\sup A$; 若集合 A 的所有下界组成的集合为 L , 则称 $\max L$ 为集合 A 的下确界, 记作 $\inf A$ 。以上确界为例, 证明中往往使用如下等价描述:

(1) 上界: $\forall x \in A, x \leq \sup A$;

(2) 最小: 比它小的都不是 A 的上界, 即 $\forall \varepsilon > 0, \exists x \in A, x > \sup A - \varepsilon$ 。

5. 确界存在定理*: 非空有上界的集合必存在唯一上确界, 非空有下界的集合必存在唯一下确界。

1.2.2 数列极限

1. 定义: 设 $\{x_n\}$ 为实数列, a 为实常数, 若对任意给定 $\varepsilon > 0$, 存在与 ε 有关的正整数 N , 对 $\forall n > N$, 均有 $|x_n - a| < \varepsilon$, 则称数列 $\{x_n\}$ 收敛于 a , 记作

$$\lim_{n \rightarrow \infty} x_n = a$$

若不存在这样的常数 a , 则称 $\{x_n\}$ 发散。

2. 无穷量: 若 $\forall G > 0, \exists N \in \mathbb{N}^*, \forall n > N, |x_n| > G$, 则称数列 $\{x_n\}$ 为无穷大量; 若 $\lim_{n \rightarrow \infty} x_n = 0$, 则称 $\{x_n\}$ 为无穷小量。

3. 极限的性质:

(1) 唯一性*: 若 $\lim_{n \rightarrow \infty} x_n = a$, 且 $\lim_{n \rightarrow \infty} x_n = b$, 则 $a = b$ 。

(2) 有界性*: 若 $\{x_n\}$ 收敛, 则 $\exists M > 0, \forall n \in \mathbb{N}^*, |x_n| < M$ 。

(3) 保序性*: 若 $\lim_{n \rightarrow \infty} x_n < \lim_{n \rightarrow \infty} y_n$, 则 $\exists N \in \mathbb{N}^*, \forall n > N, x_n < y_n$ 。

(4) 夹逼准则*: 若 $\lim_{n \rightarrow \infty} x_n = \lim_{n \rightarrow \infty} z_n = a$, 且 $\exists N \in \mathbb{N}^*, \forall n > N, x_n \leq y_n \leq z_n$, 则 $\lim_{n \rightarrow \infty} y_n = a$ 。非常重要!!

4. Stolz 定理*: 若 $\{y_n\}$ 是严格单调增加的正无穷大量, 且

$$\lim_{n \rightarrow \infty} \frac{x_n - x_{n-1}}{y_n - y_{n-1}} = a$$

其中 a 可为实常数或无穷大量, 则 $\lim_{n \rightarrow \infty} \frac{x_n}{y_n} = a$ 。

5. 收敛判定准则:

(1) 单调有界准则*: 若 $\{x_n\}$ 单调递增且有上界, 则 $\{x_n\}$ 收敛; 若 $\{x_n\}$ 单调递减且有下界, 则 $\{x_n\}$ 收敛。

(2) 柯西收敛准则*: 数列 $\{x_n\}$ 收敛的充要条件是: $\forall \varepsilon > 0, \exists N \in \mathbb{N}^*, \forall n, m > N, |x_n - x_m| < \varepsilon$ 。

6. 实数系的完备性: 由实数构成的柯西列 $\{x_n\}$ 必存在实数极限。

7. 实数系基本定理: 之前介绍过确界存在性定理、单调有界准则、柯西收敛准则。

(1) 闭区间套定理*: 若一系列闭区间 $\{[a_n, b_n]\}$ 满足 $[a_{n+1}, b_{n+1}] \subset [a_n, b_n]$, 且 $\lim_{n \rightarrow \infty} (b_n - a_n) = 0$, 则存在唯一 ξ 属于所有 $[a_n, b_n]$, 且 $\lim_{n \rightarrow \infty} a_n = \lim_{n \rightarrow \infty} b_n = \xi$ 。

(2) 波尔查诺-维尔斯特拉斯定理*: 有界数列必有收敛子列。

1.2.3 函数极限

1. 定义：设函数 $y = f(x)$ 在 x_0 点的某去心邻域内有定义， A 为实常数，若对任意给定的 $\varepsilon > 0$ ，存在与 ε 有关的 $\delta > 0$ ，使得任意 $0 < |x - x_0| < \delta$ ，均有 $|f(x) - A| < \varepsilon$ ，则称 A 是函数 $f(x)$ 在点 x_0 处的极限，记作 $\lim_{x \rightarrow x_0} f(x) = A$ 。

2. 性质：唯一性、局部有界性、局部保序性、夹逼准则，基本和数列极限一致。

3. 海涅定理*： $\lim_{x \rightarrow x_0} f(x) = A$ 的充分必要条件是：对任意满足 $\lim_{n \rightarrow \infty} x_n = x_0$ 的数列 $\{x_n\}$ ，有 $\lim_{n \rightarrow \infty} f(x_n) = A$ 。

4. 单侧极限：在函数极限定义中，若范围改为 $0 < x - x_0 < \delta$ ，则称为右侧极限；若范围改为 $0 < x_0 - x < \delta$ ，则称为左侧极限。

5. 柯西收敛准则：函数极限 $\lim_{x \rightarrow +\infty} f(x)$ 存在的充要条件是： $\forall \varepsilon > 0, \exists X > 0, \forall x_1, x_2 > X, |f(x_1) - f(x_2)| < \varepsilon$ 。

6. 无穷量：若 $\lim_{x \rightarrow x_0} f(x) = 0$ ，则称 $x \rightarrow x_0$ 时 $f(x)$ 是无穷小量；若 $\lim_{x \rightarrow x_0} f(x) = \infty$ ，则称 $x \rightarrow x_0$ 时 $f(x)$ 是无穷大量。

7. 无穷量的比较：以无穷小量为例，考虑两个无穷小量 $f(x)$ 和 $g(x)$ 的比值

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = k$$

(1) $k = 0$ ： $f(x)$ 是 $g(x)$ 的高阶无穷小量，记作 $f(x) = o(g(x))$ ；

(2) $k = \infty$ ： $f(x)$ 是 $g(x)$ 的低阶无穷小量；

(3) k 为非零实数： $f(x)$ 是 $g(x)$ 的同阶无穷小量；

(4) $k = 1$ ： $f(x)$ 与 $g(x)$ 互为等价无穷小量。

1.3 连续函数

相比于实分析、泛函分析等后继课程，数学分析的一大特点就是研究性质比较“好”的函数，比如在特定区间上连续的函数，或者可求任意阶导数的函数。诸如连续、可导、可积这种描述函数“好不好”的性质，通常称作函数分析性质。

在闭区间上连续的函数，性质往往比较“好”，这就意味着能从连续性出发，推导出这个函数满足的其它特性。因此，在研究连续函数的性质时，闭区间上的连续函数性质是非常重要的一环。

那还有没有比连续函数性质更“好”的函数呢？事实上，函数的连续性可分为“点态连续”和“一致连续”，而通常闭区间上的连续函数要求的仅仅是逐点连续。一致连续的要求更加严苛，它要求所取的 ε 必须和区间上连续点的具体位置无关。自然，一致连续函数也有更多的优良性质。

1.3.1 点态连续

1. $f(x)$ 在某点 x_0 处连续的定义：若 $f(x)$ 在 x_0 的某邻域内有定义，且 $\lim_{x \rightarrow x_0} f(x) = f(x_0)$ ，则称函数 $f(x)$ 在点 x_0 处连续，此时 x_0 是函数 $f(x)$ 的一个连续点。

2. 单侧连续：若 $f(x)$ 在 x_0 点的左极限为 $f(x_0)$ ，则 $f(x)$ 在 x_0 处左连续；若 $f(x)$ 在 x_0 点的右极限为 $f(x_0)$ ，则 $f(x)$ 在 x_0 处右连续。

3. 开区间上的连续性：若 $f(x)$ 在开区间 (a, b) 的任意一点处连续，则称 $f(x)$ 在 (a, b) 上连续。

4. 闭区间上的连续性：若 $f(x)$ 在开区间 (a, b) 上连续，且在 a 点右连续，在 b 点左连续，则称 $f(x)$ 在 $[a, b]$ 上连续。

5. 性质:

(1) 四则运算连续性: 连续函数经过四则运算后, 除去分母为零的点, 仍为连续函数;

(2) 反函数连续性*: 若函数 $f(x)$ 在闭区间上连续且严格单增, 则反函数 $f^{-1}(y)$ 同样连续且严格单增。

(3) 复合函数连续性: 两个连续函数的复合仍然为连续函数。

(4) 初等函数连续性: 所有初等函数在定义区间上均为连续函数。

6. 间断点: 若 $f(x)$ 在点 x_0 处不连续, 则称 x_0 为 $f(x)$ 的一个间断点。根据 x_0 点处的左右极限, 大致分为三类:

(1) 可去间断点: 左极限等于右极限, 但不等于 x_0 点处的函数值;

(2) 跳跃间断点: 左极限不等于右极限;

(3) 第二类间断点: 左极限和右极限至少有一个不存在。

1.3.2 闭区间上的连续函数

1. 有界性定理*: 若函数 $f(x)$ 在闭区间 $[a, b]$ 上连续, 则 $f(x)$ 在 $[a, b]$ 上有界。

2. 最值定理*: 若函数 $f(x)$ 在闭区间 $[a, b]$ 上连续, 则 $f(x)$ 在 $[a, b]$ 上必能取到最大值和最小值。

3. 零点存在性定理*: 若函数 $f(x)$ 在闭区间 $[a, b]$ 上连续, 且 $f(a) \cdot f(b) < 0$, 则存在 $\xi \in (a, b)$, 使得 $f(\xi) = 0$ 。

4. 介值定理*: 若 $f(x)$ 在闭区间 $[a, b]$ 上连续, 且最大值为 M , 最小值为 m , 则对任意 $C \in [m, M]$, 存在 $\xi \in [a, b]$, 使得 $f(\xi) = C$ 。

1.3.3 一致连续

1. 定义：设函数 $f(x)$ 在区间 X 上有定义，若 $\forall \varepsilon > 0, \exists \delta > 0, \forall x_1, x_2 \in X$ 满足 $|x_1 - x_2| < \delta$ ，均有 $|f(x_1) - f(x_2)| < \varepsilon$ 成立，则 $f(x)$ 在区间 X 上一致连续。

2. 证明非一致连续的技巧：若区间 X 上存在 $\{x_n\}$ 和 $\{y_n\}$ 满足 $\lim_{n \rightarrow +\infty} (x_n - y_n) = 0$ ，但 $\lim_{n \rightarrow +\infty} f(x_n) \neq \lim_{n \rightarrow +\infty} f(y_n)$ ，则可判定 $f(x)$ 在区间 X 上非一致连续。

3. **Cantor 定理**：若函数 $f(x)$ 在闭区间 $[a, b]$ 上连续，则 $f(x)$ 在 $[a, b]$ 上一致连续。

1.4 一元函数微分学

在做好了关于实数系、极限、连续的铺垫后，接下来将正式进入一元函数微分学的学习。

高中阶段简要介绍了导数的基本求法，以及一些基本初等函数的导数，在应用层面仅仅停留在利用导数求极限部分。事实上，导数的魅力不止于此：拉格朗日中值定理揭示了弦和切线的重要联系，是数学证明不可或缺的重要工具；洛必达法则为比较无穷小量省去诸多繁琐的推导，仅需不断地求导，便可轻松解决不定式的问题；泰勒展开作为数值逼近的手段，在计算数学中有着极为广泛的应用。

提醒一点，虽然之前已经掌握了基本的求导方法，但求导法则可不能乱用，导数的应用必须建立在函数可导的基础上。因此学习微分学的知识时，首先需要学习从定义层面上判定函数是否可导。

1.4.1 微分和导数的概念

1. **微分**: 若自变量 x 的增量 $\Delta x \rightarrow 0$ 时, 因变量 y 的增量 $\Delta y = f(x+\Delta x) - f(x)$ 可表示为

$$\Delta y = g(x_0)\Delta x + o(\Delta x)$$

则称 $f(x)$ 在 x_0 点可微; 若 $f(x)$ 在区间 X 内任一点均可微, 则称 $f(x)$ 在区间 X 上可微, 记作 $dy = g(x)dx$ 。

2. **导数**: 若 x_0 处极限

$$\lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x) - f(x_0)}{\Delta x}$$

存在, 则称 $f(x)$ 在 x_0 处可导, 并称该极限值为 $f(x)$ 在 x_0 处的导数, 记作 $f'(x_0)$; 若 $f(x)$ 在区间 X 内任一点均可导, 则称 $f(x)$ 在区间 X 上可导, 并将导函数记作 $f'(x)$ 或 $\frac{dy}{dx}$ 。

3. **可微、可导和连续的关系**: 可微与可导等价, 可导函数必连续, 连续函数不一定可导。

4. **单侧导数**: 在导数定义式中, 若取左极限, 得到的便是左导数; 若取右极限, 得到的便是右导数。

5. **函数可导的判定**: 函数在某点可导的充要条件是该点处函数的左、右导数存在且相等。

6. 常用函数及其导数:

(1) 常数函数: $(C)' = 0$;

(2) 指数函数、对数函数: $(a^x)' = a^x \ln a$, $(\log_a x)' = \frac{1}{x \ln a}$;

(3) 幂函数: $(x^a)' = ax^{a-1}$;

(4) 三角函数: $(\sin x)' = \cos x$; $(\cos x)' = -\sin x$ 。

7. 高阶微分: $d^n y = f^{(n)}(x)dx^n$, 其中 $f^{(n)}(x)$ 为对 $f(x)$ 连续求 n 次导数的结果, 又记作 $f^{(n)}(x) = \frac{d^n y}{dx^n}$ 。

1.4.2 求导法则

1. 四则运算:

(1) 加法和减法: $(f \pm g)' = f' \pm g'$;

(2) 乘法: $(f \cdot g)' = f' \cdot g + f \cdot g'$;

(3) 除法: $\left(\frac{f}{g}\right)' = \frac{f' \cdot g - f \cdot g'}{g^2}$ 。

2. 反函数求导法则: 设反函数为 $x = f(y)$, 对两边求微分得 $dx = f'(y)dy$, 整理得 $\frac{dy}{dx} = \frac{1}{f'(y)}$ 。

3. 复合函数求导法则: 设 $y = f(g(x))$ 是 $y = f(u)$ 和 $u = g(x)$ 两个函数的复合, 则 $y'(x) = y'(u) \cdot u'(x)$, 或记作

$$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$$

4. 一阶微分的形式不变性: 对函数 $y = f(x)$ 而言, 无论 x 是因变量还是中间变量, 一阶微分的形式总是相同的:

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

5. 隐函数求导法则: 设隐函数 $F(x, y) = 0$, 将 y 视为关于 x 的函数 $y(x)$, 对两边关于 x 求导, 整理出 y' 的显式函数表达式。

6. 参数方程求导法则: 对参数方程 $x = \varphi(t)$, $y = \psi(t)$, 先求 $\frac{dy}{dt} = \psi'(t)$,

$\frac{dy}{dt} = \psi'(t)$, 再将两者相除, 消去 dt , 得到

$$\frac{dy}{dx} = \frac{\psi'(t)}{\varphi'(t)}$$

1.4.3 微分中值定理

1. **极值**: 设 $f(x)$ 在 (a, b) 上有定义, $x_0 \in (a, b)$, 若存在 x_0 的某一邻域, 使得邻域内任意一点 x , 均有 $f(x) \leq f(x_0)$, 则称 x_0 为 $f(x)$ 的一个极大值点; $f(x) \geq f(x_0)$ 时, 称 x_0 为 $f(x)$ 的一个极小值点。

2. **费马引理***: 设 x_0 是 $f(x)$ 的一个极值点, 且 $f(x)$ 在 x_0 处可导, 则 $f'(x_0) = 0$ 。

3. **罗尔中值定理***: 设 $f(x)$ 在 $[a, b]$ 上连续, 在 (a, b) 上可导, $f(a) = f(b)$, 则存在 $\xi \in (a, b)$, 使得 $f'(\xi) = 0$ 。

4. **拉格朗日中值定理***: 设 $f(x)$ 在 $[a, b]$ 上连续, 在 (a, b) 上可导, 则存在 $\xi \in (a, b)$, 使得

$$f'(\xi) = \frac{f(b) - f(a)}{b - a}$$

5. **柯西中值定理***: 设 $f(x)$ 和 $g(x)$ 在 $[a, b]$ 上连续, 在 (a, b) 上可导, 且 $g'(x) \neq 0$, 则存在 $\xi \in (a, b)$, 使得

$$\frac{f'(\xi)}{g'(\xi)} = \frac{f(b) - f(a)}{g(b) - g(a)}$$

6. **洛必达法则***: 若 $\lim_{x \rightarrow x_0} f(x)$ 和 $\lim_{x \rightarrow x_0} g(x)$ 同为 0 或 ∞ , 则有

$$\lim_{x \rightarrow x_0} \frac{f(x)}{g(x)} = \lim_{x \rightarrow x_0} \frac{f'(x)}{g'(x)}$$

1.4.4 泰勒展开

1. 带 **Peano** 余项的泰勒展开*: 设 $f(x)$ 在 x_0 处有 n 阶导数, 则存在 x_0 的一个邻域, 对该邻域内任意一点 x , 成立

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + o((x - x_0)^n)$$

2. 带 **Lagrange** 余项的泰勒展开: 条件同上, ξ 为 x 和 x_0 之间一点, 成立

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + \frac{f^{(n+1)}(\xi)}{(n+1)!}(x - x_0)^{n+1}$$

3. 麦克劳林公式: 取 $x_0 = 0$, 其中 $r_n(x)$ 为皮亚诺余项或拉格朗日余项:

$$f(x) = f(0) + f'(0)x + \frac{f''(0)}{2!}x^2 + \cdots + \frac{f^{(n)}(0)}{n!}x^n + r_n(x)$$

4. 常用函数的麦克劳林展开:

$$(1) e^x = 1 + x + \frac{x^2}{2!} + \cdots + \frac{x^n}{n!} + r_n(x);$$

$$(2) \ln(1+x) = x - \frac{x^2}{2!} + \frac{x^3}{3!} \cdots + \frac{(-1)^{n+1}x^n}{n!} + r_n(x);$$

$$(3) \sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} + \cdots + (-1)^n \frac{x^{2n+1}}{(2n+1)!} + r_{2n+2}(x);$$

$$(4) \cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} + \cdots + (-1)^n \frac{x^{2n}}{(2n)!} + r_{2n+1}(x)。$$

5. 应用举例: 求未定式极限; 证明和导数相关的不等式; 求曲线的渐近线方程。

1.5 一元函数积分学

不定积分是和微分相对应的概念。就和乘法、除法的关系一样，不定积分就是把微分作逆运算，倒回去求原函数。定积分则是用于求不规则体面积的工具，将横轴分为若干小段，在每一段内求矩形面积再求和。

这两个出发点完全不同的概念，仅仅用了一个公式就紧密地联系在一起，这就是一元函数微积分中最伟大的结论之一：Newton-Leibniz 公式。有了这个公式后，定积分和不定积分的求解方法便可完全互通，知道不定积分和上下限，便能轻松求出不规则区域的面积。

这个公式的意义不止于此，若将积分上下限改为间断点或无穷远处，仍可以通过微积分基本定理转化为不定积分求解问题。这种积分限较为特殊的积分称为广义积分，除了关心广义积分的具体值外，往往还会研究广义积分的收敛性，并引申出很多收敛性的判别法。

1.5.1 不定积分

1. **原函数**：若函数 $F(x)$ 和 $f(x)$ 满足 $F'(x) = f(x)$ ，则称 $F(x)$ 为 $f(x)$ 的原函数。

2. **不定积分**：函数 $f(x)$ 原函数的全体称为 $f(x)$ 的不定积分，记作 $\int f(x)dx$ 。若 $f(x)$ 的某一个原函数为 $F(x)$ ，则 $\int f(x)dx = F(x) + C$ ，其中 C 为任意常数。

3. **第一类换元积分法**：若 $f(x)$ 可等价变化为 $f(u(x)) \cdot u'(x)$ ，则用 $du = u'(x)dx$ 代换，得

$$\int f(x)dx = \int f(u)du$$

即用中间变量 u 代替原本的变量 x 进行积分。当前的积分变元由微分是 dx 还是 du 决定，下同。

4. 第二类换元积分法：代入 $dx = x'(u)du$ ，得

$$\int f(x)dx = \int f(x(u)) \cdot x'(u)du$$

即把原本变量 x 替换为中间变量 u 积分。

5. 分部积分法：利用微分的乘法法则 $d(u \cdot v) = u \cdot dv + v \cdot du$ ，两边求积分得

$$u \cdot v = \int u dv + \int v du$$

因此，在求 $\int f(x)dg(x)$ 时，可以借助分部积分，转化为 $f(x) \cdot g(x) - \int g(x)df(x)$ ，交换待积式和微分式。

1.5.2 定积分

1. 定义：设 $f(x)$ 是定义在 $[a, b]$ 上的有界函数，在 $[a, b]$ 上任意取划分 P ：
 $a = x_0 < x_1 < \cdots < x_n = b$ ，在区间 $[x_{i-1}, x_i]$ 内任取一点 ξ_i 。记小区间 $[x_{i-1}, x_i]$ 的长度为 $\Delta x_i = x_i - x_{i-1}$ ，并记小区间长度的最大值为 $\lambda = \max_{1 \leq i \leq n} \Delta x_i$ 。若如下极限存在：

$$\lim_{\lambda \rightarrow 0} \sum_{i=1}^n f(\xi_i) \Delta x_i$$

且极限值与划分 P 和点 ξ 的取法均无关，则称该极限值为 $f(x)$ 在 $[a, b]$ 上的定积分，记作 $\int_a^b f(x)dx$ ，此时称 $f(x)$ 在 $[a, b]$ 上可积。

2. 可积性条件：

(1) 达布和：令 $M_i = \max_{x_{i-1} \leq x \leq x_i} f(x)$ ， $m_i = \min_{x_{i-1} \leq x \leq x_i} f(x)$ ，则称 $\sum_{i=1}^n M_i \Delta x_i$ 为达布大和， $\sum_{i=1}^n m_i \Delta x_i$ 为达布小和。

(2) 黎曼可积的充要条件： $f(x)$ 在 $[a, b]$ 上可积的充要条件是：对任意

划分 P , 当区间长度的最大值 $\lambda \rightarrow 0$ 时, 达布大和与达布小和的极限值相等。

(3) 振幅与可积性条件: 记振幅 $\omega_i = M_i - m_i$, 则有界函数 $f(x)$ 对任意划分 P , 当区间长度的最大值 $\lambda \rightarrow 0$ 时, $\lim_{\lambda \rightarrow 0} \sum_{i=1}^n \omega_i \Delta x_i = 0$ 。

(4) 闭区间上函数的可积性*: 闭区间上的连续函数必定可积; 闭区间上的单调函数必定可积。

(5) 可积的常用判定方法*: 有界函数 $f(x)$ 在 $[a, b]$ 上可积的充要条件是: 对任意给定的 $\varepsilon > 0$, 存在一种划分 P , 使得 $\sum_{i=1}^n \omega_i \Delta x_i < \varepsilon$ 。

3. 定积分的性质:

(1) 线性性质: $\int_a^b (k_1 f(x) + k_2 g(x)) dx = k_1 \int_a^b f(x) dx + k_2 \int_a^b g(x) dx$;

(2) 保序性*: 若 $[a, b]$ 上恒有 $f(x) \geq g(x)$, 则 $\int_a^b f(x) dx \geq \int_a^b g(x) dx$ 。

(3) 绝对可积性*: 若 $f(x)$ 在 $[a, b]$ 上可积, 则 $|f(x)|$ 在 $[a, b]$ 上也可积, 且成立

$$\left| \int_a^b f(x) dx \right| \leq \int_a^b |f(x)| dx$$

(4) 区间可加性*: 若 $c \in [a, b]$, 则 $f(x)$ 在 $[a, b]$ 上可积的充要条件是 $f(x)$ 在 $[a, c]$ 和 $[c, b]$ 上均可积, 且满足

$$\int_a^b f(x) dx = \int_a^c f(x) dx + \int_c^b f(x) dx$$

4. 积分第一中值定理*:

(1) 若 $f(x)$ 和 $g(x)$ 在区间 $[a, b]$ 上可积, $g(x)$ 在 $[a, b]$ 上不变号, 记 $\sup_{[a, b]} f(x) = M$, $\inf_{[a, b]} f(x) = m$, 则存在 $\eta \in [m, M]$, 满足

$$\int_a^b f(x) g(x) dx = \eta \int_a^b g(x) dx$$

(2) 在上述条件下, 若 $f(x)$ 在 $[a, b]$ 上连续, 则存在 $\xi \in [a, b]$, 使得

$$\int_a^b f(x)g(x)dx = f(\xi) \int_a^b g(x)dx$$

5. 变限积分: 设 $f(x)$ 在 $[a, b]$ 上连续, 作函数 $F(x) = \int_a^x f(t)dt$, 其中 $x \in [a, b]$, 则该函数称为 $f(x)$ 的变上限积分, 同理定义 $F(x) = \int_x^b f(t)dt$ 为 $f(x)$ 的变下限积分。此时 $F(x)$ 在 $[a, b]$ 上为可微函数, 且 $F'(x) = f(x)$ 。

6. 微积分基本定理: 设 $f(x)$ 在 $[a, b]$ 上连续, $F(x)$ 是 $f(x)$ 在 $[a, b]$ 上的一个原函数, 则有

$$\int_a^b f(x)dx = F(b) - F(a)$$

此公式便是著名的牛顿-莱布尼茨公式。

7. 定积分用于几何计算:

(1) $f(x)$ 与 x 轴围成的平面图形面积: 设图形所处横坐标区间为 $[a, b]$, 则该面积为 $\int_a^b f(x)dx$;

(2) $f(x)$ 图像上一段曲线的弧长: 设曲线所处横坐标区间为 $[a, b]$, 则曲线弧长为 $\int_a^b \sqrt{1 + [f'(x)]^2}dx$;

(3) $f(x)$ 围绕 x 轴旋转一圈形成的旋转体体积: 设旋转体所处横坐标区间为 $[a, b]$, 则旋转体体积为 $\pi \int_a^b [f(x)]^2 dx$;

(4) 极坐标方程 $r = r(\theta)$ 下平面图形面积: 设图形所处角度区间为 $[\alpha, \beta]$, 则面积为 $\frac{1}{2} \int_\alpha^\beta r^2(\theta)d\theta$ 。

1.5.3 广义积分

1. 定义: 积分区间无限或者被积函数无界的定积分。

2. 收敛性: 设函数 $f(x)$ 在 $[a, +\infty]$ 上有定义, 且在任意有限区间 $[a, A]$ 上可

积。若极限

$$\lim_{A \rightarrow +\infty} \int_a^A f(x) dx$$

存在, 则称广义积分 $\int_a^{+\infty} f(x) dx$ 收敛, 积分值等于上述极限值; 若该极限不存在, 则称广义积分 $\int_a^{+\infty} f(x) dx$ 发散。

3. 柯西主值: 若

$$\lim_{A \rightarrow +\infty} \int_{-A}^A f(x) dx = \lim_{A \rightarrow +\infty} [F(A) - F(-A)]$$

收敛, 则称该极限为 $\int_{-\infty}^{+\infty} f(x) dx$ 的柯西主值, 记作 $(\text{cpv}) \int_{-\infty}^{+\infty} f(x) dx$ 。

4. 柯西收敛准则: 广义积分 $\int_a^{+\infty} f(x) dx$ 收敛的充分必要条件是: $\forall \varepsilon > 0$, $\exists N \geq a$, $\forall A_1, A_2 > N$, 有

$$\left| \int_{A_1}^{A_2} f(x) dx \right| < \varepsilon$$

5. 绝对收敛: 若广义积分 $\int_a^{+\infty} |f(x)| dx$ 收敛, 则称 $\int_a^{+\infty} f(x) dx$ 绝对收敛; 若 $\int_a^{+\infty} |f(x)| dx$ 发散而 $\int_a^{+\infty} f(x) dx$ 收敛, 则称 $\int_a^{+\infty} f(x) dx$ 条件收敛。

6. 非负函数广义积分的收敛判别法:

(1) 比较判别法: 若 $[a, +\infty)$ 上恒有 $0 \leq f(x) \leq K\varphi(x)$, 则 $\int_a^{+\infty} \varphi(x) dx$ 收敛时 $\int_a^{+\infty} f(x) dx$ 收敛, $\int_a^{+\infty} f(x) dx$ 发散时 $\int_a^{+\infty} \varphi(x) dx$ 发散。

(2) 比较判别法的极限形式: 考虑 $\lim_{x \rightarrow +\infty} \frac{f(x)}{\varphi(x)} = l$, 当 l 为实常数时 $\int_a^{+\infty} f(x) dx$ 和 $\int_a^{+\infty} \varphi(x) dx$ 同时收敛或发散; 当 $l = \infty$ 时, $\int_a^{+\infty} \varphi(x) dx$ 发散意味着 $\int_a^{+\infty} f(x) dx$ 发散; 当 $l = 0$ 时, $\int_a^{+\infty} \varphi(x) dx$ 收敛意味着 $\int_a^{+\infty} f(x) dx$ 收敛。

(3) 柯西判别法: 若 $f(x) \leq \frac{K}{x^p}$, $p > 1$, 则 $\int_a^{+\infty} f(x) dx$ 收敛; 若 $f(x) \geq \frac{K}{x^p}$, $p \leq 1$, 则 $\int_a^{+\infty} f(x) dx$ 发散。

7. 积分第二中值定理: 设 $f(x)$ 在 $[a, b]$ 上可积, $g(x)$ 在 $[a, b]$ 上单调, 则存

在 $\xi \in [a, b]$, 使得

$$\int_a^b f(x)g(x)dx = g(a) \int_a^\xi f(x)dx + g(b) \int_\xi^b f(x)dx$$

8. Abel-Dirichlet 判别法: 若下面某个条件满足, 则广义积分 $\int_a^{+\infty} f(x)g(x)dx$ 收敛:

(1) Abel 条件: $\int_a^{+\infty} f(x)dx$ 收敛, $g(x)$ 在 $[a, +\infty)$ 上单调有界;

(2) Dirichlet 条件: $F(A) = \int_a^A f(x)dx$ 在 $A \in [a, +\infty)$ 上有界, $g(x)$ 在 $[a, +\infty)$ 上单调且 $\lim_{x \rightarrow +\infty} g(x) = 0$ 。

1.6 无穷级数

无穷级数研究的目的是无穷项的和。当求和的每一项都是确定的数时, 该级数被称作数项级数; 当求和的项和某一自变量 x 相关时, 求和的结果会是关于 x 的函数, 被称作函数项级数。

数项级数分为两种情况: 当求和结果随着项数增加而越来越逼近某一特定值时, 该级数收敛; 当求和结果无法控制在某一具体值附近时, 该级数发散。因此, 研究常数项级数的性质时, 不仅要掌握常数项级数的求法, 还需要掌握判断级数敛散性的方法。

函数项级数也分为两种情况: 点态收敛和一致收敛。在 x 的定义域内, 若任取一点 x_0 代入, 得到的数项级数都收敛, 则函数项级数是点态收敛的; 一致收敛的条件更为苛刻, 性质也更为优异。

傅里叶级数可以将任意信号转化为正弦、余弦信号的叠加, 由此衍生出的傅里叶变换、离散余弦变换、快速傅里叶变换提供了时域-频域转化的方法, 是音频和图像信号处理的一大利器。

1.6.1 数项级数

1. 定义：记部分和数列 $S_n = \sum_{i=1}^n x_i$ ，若 $\{S_n\}$ 收敛于有限数 S ，则称无穷级数 $\sum_{n=1}^{\infty} x_n$ 收敛，记作

$$\sum_{n=1}^{\infty} x_n = S$$

若部分和数列 $\{S_n\}$ 发散，则称无穷级数 $\sum_{n=1}^{\infty} x_n$ 发散。

2. 性质*：若 $\sum_{n=1}^{\infty} x_n$ 收敛，则必有 $\lim_{n \rightarrow \infty} x_n = 0$ 。

3. 正项级数：

(1) 定义：每一项均为非负实数的级数，即 $\forall n \in \mathbb{N}$, $x_n \geq 0$ ，则称 $\sum_{n=1}^{\infty} x_n$ 为正项级数。

(2) 收敛原理：正项级数收敛的充要条件是它的部分和数列有上界。

(3) 比较判别法*：若存在 $N \in \mathbb{N}^*$ 和常数 $A > 0$ ，使得 $n > N$ 时恒有 $x_n \leq A y_n$ ，则 $\sum_{n=1}^{\infty} y_n$ 收敛意味着 $\sum_{n=1}^{\infty} x_n$ 收敛， $\sum_{n=1}^{\infty} x_n$ 发散意味着 $\sum_{n=1}^{\infty} y_n$ 发散。

(4) 比较判别法的极限形式：设 $\lim_{n \rightarrow \infty} \frac{x_n}{y_n} = l$ ，若 l 为实常数，则 $\sum_{n=1}^{\infty} x_n$ 和 $\sum_{n=1}^{\infty} y_n$ 同敛散性；若 $l = 0$ ，则 $\sum_{n=1}^{\infty} y_n$ 收敛意味着 $\sum_{n=1}^{\infty} x_n$ 收敛；若 $l = \infty$ ，则 $\sum_{n=1}^{\infty} y_n$ 发散意味着 $\sum_{n=1}^{\infty} x_n$ 发散。

(5) 柯西判别法*：考虑 $r = \lim_{n \rightarrow \infty} \sqrt[n]{x_n}$ 。若 $r > 1$ ，则 $\sum_{n=1}^{\infty} x_n$ 发散；若 $r < 1$ ，则 $\sum_{n=1}^{\infty} x_n$ 收敛；若 $r = 1$ ，则柯西判别法失效。

(6) 达朗贝尔判别法*：考虑 $r = \lim_{n \rightarrow \infty} \frac{x_{n+1}}{x_n}$ 。若 $r > 1$ ，则 $\sum_{n=1}^{\infty} x_n$ 发散；若 $r < 1$ ，则 $\sum_{n=1}^{\infty} x_n$ 收敛；若 $r = 1$ ，则达朗贝尔判别法失效。

(7) 拉比判别法：考虑 $r = \lim_{n \rightarrow \infty} n \left(\frac{x_n}{x_{n+1}} \right)$ 。若 $r > 1$ ，则级数 $\sum_{n=1}^{\infty} x_n$ 收敛；若

$r < 1$, 则级数 $\sum_{n=1}^{\infty} x_n$ 发散; 若 $r = 1$, 则拉比判别法失效。

(8) 广义积分判别法: 正项级数 $\sum_{n=1}^{\infty} \left[\int_{a_n}^{a_{n+1}} f(x) dx \right]$ 与广义积分 $\int_a^{+\infty} f(x) dx$ 敛散性一致, 其中 $\lim_{n \rightarrow \infty} a_n = +\infty$ 。

4. 任意项级数:

(1) 柯西收敛准则: 级数 $\sum_{n=1}^{\infty} x_n$ 收敛的充要条件是: $\forall \varepsilon > 0, \exists N \in \mathbb{N}^*, \forall m > n > N, \left| \sum_{k=n}^m x_k \right| < \varepsilon$ 。

(2) 莱布尼茨判别法*: 若正项级数 $\{x_n\}$ 单调递减且收敛于 0, 则交错级数 $\sum_{n=1}^{\infty} (-1)^n x_n$ 收敛。

(3) Abel-Dirichlet 判别法: 若下面某个条件满足, 则无穷级数 $\sum_{n=1}^{\infty} a_n b_n$ 收敛:

· Abel 条件: $\sum_{n=1}^{\infty} b_n$ 收敛, $\{a_n\}$ 单调有界;

· Dirichlet 条件: $\sum_{n=1}^{\infty} b_n$ 有界, $\{a_n\}$ 单调趋于 0。

(4) 绝对收敛: 若级数 $\sum_{n=1}^{\infty} |x_n|$ 收敛, 则称 $\sum_{n=1}^{\infty} x_n$ 绝对收敛; 若 $\sum_{n=1}^{\infty} |x_n|$ 发散, 且 $\sum_{n=1}^{\infty} x_n$ 收敛, 则称 $\sum_{n=1}^{\infty} x_n$ 条件收敛。

1.6.2 函数项级数

1. 定义: 设 $u_n(x)$ 是具有公共定义域的一系列函数, 对无穷个函数求和的结果 $\sum_{n=1}^{\infty} u_n(x)$ 称为函数项级数。

2. 点态收敛性: 若对定义域内给定一点 x_0 , 数项级数 $\sum_{n=1}^{\infty} u_n(x_0)$ 收敛, 则称 $\sum_{n=1}^{\infty} u_n(x)$ 在点 x_0 处收敛。 $\sum_{n=1}^{\infty} u_n(x)$ 所有收敛点的全体称为 $\sum_{n=1}^{\infty} u_n(x)$ 的收敛域。

3. 和函数: 若 $\sum_{n=1}^{\infty} u_n(x)$ 的收敛域为 D , 则 $\sum_{n=1}^{\infty} u_n(x)$ 在 D 上定义了一个关于 x 的函数 $S(x) = \sum_{n=1}^{\infty} u_n(x)$, 称作 $\sum_{n=1}^{\infty} u_n(x)$ 的和函数。

4. 一致收敛性:

(1) 定义: 若对任意给定的 $\varepsilon > 0$, 存在仅与 ε 有关而与 x 无关的 $N \in \mathbb{N}^*$, 当 $n > N$ 时,

$$|S_n(x) - S(x)| < \varepsilon$$

则称 $\{S_n(x)\}$ 一致收敛于函数 $S(x)$ 。

(2) 内闭一致收敛: 若对任意闭区间 $[a, b] \subset D$, 均有 $\{S_n(x)\}$ 在 $[a, b]$ 上一致收敛于 $S(x)$, 则称 $\{S_n(x)\}$ 在 D 上内闭一致收敛于 $S(x)$ 。

(3) 判定定理: 设 $\{S_n(x)\}$ 在 D 上点态收敛于 $S(x)$, 定义 $\{S_n(x)\}$ 与 $S(x)$ 的距离为 $d(S_n, S) = \sup_{x \in D} |S_n(x) - S(x)|$, 则 $\{S_n(x)\}$ 在 D 上一致收敛于 $S(x)$ 的充分必要条件是:

$$\lim_{n \rightarrow \infty} d(S_n, S) = 0$$

(4) 非一致收敛判定: 若存在数列 $\{x_n\}$, $x_n \in D$, 使 $\lim_{n \rightarrow \infty} (S_n(x_n) - S(x_n)) \neq 0$, 则 $\{S_n(x)\}$ 在 D 上不一致收敛于 $S(x)$ 。

5. 一致收敛级数:

(1) 定义: 记 $S_n(x)$ 为 $\{u_n(x)\}$ 的部分和数列, 若 $S_n(x)$ 一致收敛于 $S(x)$, 则函数项级数 $\sum_{n=1}^{\infty} u_n(x)$ 一致收敛于 $S(x)$ 。

(2) 柯西收敛准则: $\sum_{n=1}^{\infty} u_n(x)$ 在 D 上一致收敛的充要条件是: $\forall \varepsilon > 0$, $\exists N \in \mathbb{N}^*$, $\forall m > n > N$, $x \in D$, $|u_{n+1}(x) + \cdots + u_m(x)| < \varepsilon$ 。

(3) 维尔斯特拉斯判别法: 若 $\forall x \in D$, $|u_n(x)| \leq a_n$, 且 $\sum_{n=1}^{\infty} a_n$ 收敛, 则 $\sum_{n=1}^{\infty} u_n(x)$ 在 D 上一致收敛。

(4) Abel-Dirichlet 判别法: 若下面某条件满足, 则函数项级数 $\sum_{n=1}^{\infty} a_n(x)b_n(x)$ 在 D 上一致收敛:

· Abel 条件: $\sum_{n=1}^{\infty} b_n(x)$ 在 D 上一致收敛, 且对任意 $x \in D$, $\{a_n(x)\}$ 关

于 n 单调且一致有界;

· Dirichlet 条件: $\sum_{n=1}^{\infty} b_n(x)$ 的部分和序列在 D 上一致有界, 且对任意 $x \in D$, $\{a_n(x)\}$ 关于 n 单调且一致收敛于 0。

(5) 性质: 可逐项积分; 可逐项求极限; 可逐项求导。

6. 幂级数:

(1) 定义: 形如 $\sum_{n=0}^{\infty} a_n x^n$ 的函数项级数称为幂级数。

(2) 收敛域: 定义收敛半径 $R = \frac{1}{\lim_{n \rightarrow \infty} \sqrt[n]{|a_n|}}$, 当 $|x| < R$ 时幂级数收敛, $|x| > R$ 时幂级数发散, $|x| = R$ 时单独代入判断。使幂级数收敛的 x 全体称为幂级数的收敛域。

(3) 达朗贝尔判别法: 若 $\lim_{n \rightarrow \infty} \left| \frac{a_{n+1}}{a_n} \right| = A$, 则幂级数的收敛半径为 $R = \frac{1}{A}$ 。

(4) 性质: 幂级数在其收敛域上连续, 在收敛域内部可逐项求导、逐项积分。

7. 泰勒级数: 当泰勒展开

$$f(x) = f(x_0) + f'(x_0)(x - x_0) + \cdots + \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n + r_n(x)$$

中, 余项满足 $\lim_{n \rightarrow \infty} r_n(x) = 0$ 时, $f(x)$ 可展开成幂级数 $\sum_{n=0}^{\infty} \frac{f^{(n)}(x_0)}{n!}(x - x_0)^n$, 且幂级数收敛于函数 $f(x)$ 。

1.6.3 傅里叶级数

1. 定义: 设 $f(x)$ 是以 2π 为周期的函数, 且在 $[-\pi, \pi]$ 上可积或绝对可积, 则 $f(x)$ 可展开为如下 Fourier 级数:

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} (a_n \cos nx + b_n \sin nx)$$

其中 $a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx dx$, $b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx dx$ 。

2. 正弦级数和余弦级数: 若 $f(x)$ 为奇函数, 则 $a_n = 0$, 即 $f(x)$ 可展开为正弦级数 $\sum_{n=1}^{\infty} b_n \sin nx$; 若 $f(x)$ 为偶函数, 则 $b_n = 0$, 即 $f(x)$ 可展开为余弦级数

$$\frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx.$$

3. 任意周期函数的 **Fourier** 展开: 设 $f(x)$ 是以 $2T$ 为周期的函数, 且在 $[-T, T]$ 上可积或绝对可积, 则 $f(x)$ 可展开为如下 Fourier 级数:

$$f(x) \sim \frac{a_0}{2} + \sum_{n=1}^{\infty} \left(a_n \cos \frac{n\pi}{T} x + b_n \sin \frac{n\pi}{T} x \right)$$

其中 $a_n = \frac{1}{T} \int_{-T}^T f(x) \cos \frac{n\pi}{T} x dx$, $b_n = \frac{1}{T} \int_{-T}^T f(x) \sin \frac{n\pi}{T} x dx$ 。

4. 单点处的收敛性: 若 $f(x)$ 在 $[-\pi, \pi]$ 上可积或绝对可积, 在点 x 处存在两个单侧导数, 则 $f(x)$ 的傅里叶级数在点 x 处收敛于 $\frac{f(x+) + f(x-)}{2}$ 。

5. 性质: 可逐项积分; 可逐项微分。

6. 傅里叶变换:

$$\mathcal{F}(f) = \int_{-\infty}^{+\infty} f(x) e^{-i\omega x} dx$$

称为 $f(x)$ 的傅里叶变换; 而

$$\mathcal{F}^{-1}(f) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} f(\omega) e^{i\omega x} d\omega$$

称为 $f(\omega)$ 的傅里叶逆变换。

7. 卷积定理: 记 $f(x)$ 和 $g(x)$ 的卷积为

$$(f * g)(x) = \int_{-\infty}^{+\infty} f(t) g(x - t) dt$$

则傅里叶变换将卷积化为乘积, 即 $\mathcal{F}[f * g] = \mathcal{F}[f] \cdot \mathcal{F}[g]$ 。

1.7 多元函数微分学

多元函数是指欧几里得空间 \mathbb{R}^n 的子集到 \mathbb{R} 上的映射。由于高维欧氏空间的部分概念如极限、连续、区间都无法直接套用实数域上的定义，因此在引入多元函数的偏导数和全微分之前，有必要重新定义欧氏空间上的部分基础概念。

偏导数就是多元函数针对某一坐标轴的方向导数，如 $\frac{\partial f}{\partial x}$ 就固定 y 坐标，只对 x 方向求导。在此基础上，定义了多元函数的全微分，并指出多元函数的可微和可偏导不是对等的。偏导数还有多种求导法则，如针对复合函数的链式法则，以及针对隐函数的逆映射定理。

偏导数的应用非常广泛。在空间解析几何中，利用偏导数可以求曲线的切线和法平面，以及曲面的切平面与法线；在最优化问题中，不仅可以求无条件限制的极值，还可用拉格朗日乘子法求解条件极值。

1.7.1 欧几里得空间

1. 定义：设 $\mathbb{R}^n = (x_1, \dots, x_n) | x_i \in \mathbb{R}$ 为 n 个 \mathbb{R} 的笛卡尔积，在此基础上定义加法运算、数乘运算、向量内积和距离，即构成欧氏空间，其中的元素称为向量。

(1) 加法运算：设 $\mathbf{x} = (x_1, \dots, x_n)$, $\mathbf{y} = (y_1, \dots, y_n)$, 定义 $\mathbf{x} + \mathbf{y} = (x_1 + y_1, \dots, x_n + y_n)$;

(2) 数乘运算：设 $\lambda \in \mathbb{R}$, $\mathbf{x} = (x_1, \dots, x_n)$, 定义 $\lambda \mathbf{x} = (\lambda x_1, \dots, \lambda x_n)$;

(3) 内积：定义 \mathbf{x} 和 \mathbf{y} 的内积 $\langle \mathbf{x}, \mathbf{y} \rangle = x_1 y_1 + \dots + x_n y_n = \sum_{i=1}^n x_i y_i$;

(4) 距离：定义 \mathbf{x} 和 \mathbf{y} 的距离 $|\mathbf{x} - \mathbf{y}| = \sqrt{(x_1 - y_1)^2 + \dots + (x_n - y_n)^2}$;

(5) 范数：定义 $\|\mathbf{x}\| = \sqrt{\langle \mathbf{x}, \mathbf{x} \rangle} = \sqrt{x_1^2 + \dots + x_n^2}$ 为 \mathbf{x} 的 Euclid 范数。

2. 内积的性质:

(1) 正定性: $\langle \mathbf{x}, \mathbf{x} \rangle \geq 0$, 当且仅当 $\mathbf{x} = \mathbf{0}$ 时取等;(2) 对称性: $\langle \mathbf{x}, \mathbf{y} \rangle = \langle \mathbf{y}, \mathbf{x} \rangle$;(3) 线性性: $\langle \lambda \mathbf{x} + \mu \mathbf{y}, \mathbf{z} \rangle = \lambda \langle \mathbf{x}, \mathbf{z} \rangle + \mu \langle \mathbf{y}, \mathbf{z} \rangle$;(4) 柯西不等式*: $\langle \mathbf{x}, \mathbf{y} \rangle^2 \leq \langle \mathbf{x}, \mathbf{x} \rangle \cdot \langle \mathbf{y}, \mathbf{y} \rangle$ 。

3. 距离的性质:

(1) 正定性: $|\mathbf{x} - \mathbf{y}| \geq 0$, 当且仅当 $\mathbf{x} = \mathbf{y}$ 取等;(2) 对称性: $|\mathbf{x} - \mathbf{y}| = |\mathbf{y} - \mathbf{x}|$;(3) 三角不等式*: $|\mathbf{x} - \mathbf{z}| \leq |\mathbf{x} - \mathbf{y}| + |\mathbf{y} - \mathbf{z}|$ 。4. 邻域: 设 $\mathbf{a} = (a_1, \dots, a_n) \in \mathbb{R}^n$, $\delta > 0$, 称点集

$$O(\mathbf{a}, \delta) = \{\mathbf{x} \in \mathbb{R}^n \mid |\mathbf{x} - \mathbf{a}| < \delta\}$$

为点 \mathbf{a} 的 δ 邻域, 记作 $O(\mathbf{a}, \delta)$ 。

5. 点和 \mathbb{R}^n 上点集的关系: 设 $\mathbf{x} \in \mathbb{R}^n$, $S \subset \mathbb{R}^n$,(1) 内点: 若存在 $\delta > 0$, 使得 $O(\mathbf{x}, \delta) \subset S$, 则称 \mathbf{x} 为 S 的内点;

(2) 边界点: 若 \mathbf{x} 的任意邻域 $O(\mathbf{x}, \delta)$ 内, 都同时存在 S 内和 S 外的点, 则称 \mathbf{x} 为 S 的边界点;

(3) 外点: 若存在 $\delta > 0$, 使得 $O(\mathbf{x}, \delta) \subset S^c$, 则称 \mathbf{x} 为 S 的外点;

(4) 聚点: 若 \mathbf{x} 的任意邻域内都含有 S 中的无穷多个点, 则称 \mathbf{x} 为 S 的聚点。

6. 开集和闭集:

(1) 开集: 若 S 中的每一个点都是 S 的内点, 则称 S 为开集;(2) 闭集: 若 S 的所有聚点均属于 S , 则称 S 为闭集。

7. 欧氏空间基本定理:

(1) 闭矩形套定理: 设 $\Delta_k = [a_k, b_k] \times [c_k, d_k]$ 是 \mathbb{R}^2 上一列矩形, 若 $\Delta_{k+1} \subset \Delta_k$, 且 $\lim_{k \rightarrow \infty} \sqrt{(b_k - a_k)^2 + (d_k - c_k)^2} = 0$, 则存在唯一的点 \mathbf{a} 在所有 Δ_k 内, 且该点的横坐标等于 $\{a_k\}$ 和 $\{b_k\}$ 的极限值, 纵坐标等于 $\{c_k\}$ 和 $\{d_k\}$ 的极限值。

(2) Bolzano-Weierstrass 定理: \mathbb{R}^n 上的有界点列必有收敛子列。

(3) 柯西收敛准则: \mathbb{R}^n 上的点列 $\{x_n\}$ 收敛的充要条件是: $\forall \varepsilon > 0, \exists N \in \mathbb{N}^*, \forall m > n > N, |\mathbf{x}_m - \mathbf{x}_n| < \varepsilon$ 。

1.7.2 多元连续函数

1. 多元函数: 设 D 是 \mathbb{R}^n 上的点集, D 到 \mathbb{R} 上的映射 $f: D \rightarrow \mathbb{R}$ 称为 n 元函数。

2. 多元函数的 n 重极限: 对任意 $\varepsilon > 0$, 存在 $\delta > 0$, 使得 \mathbf{x}_0 的去心邻域 $O(\mathbf{x}_0, \delta) \setminus \{\mathbf{x}_0\}$ 内的任一点 \mathbf{x} , 均有 $|f(\mathbf{x}) - A| < \varepsilon$, 则称多元函数 $f(\mathbf{x})$ 在 \mathbf{x}_0 处收敛于 A , 记作 $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = A$ 。

3. 累次极限: 对二元函数 $f(x, y)$ 而言, 若先求 $x \rightarrow x_0$ 时 $f(x, y)$ 的极限, 再求 $y \rightarrow y_0$ 时的极限, 得到的结果称为先 x 后 y 的二次极限, 记作 $\lim_{y \rightarrow y_0} \lim_{x \rightarrow x_0} f(x, y)$; 同理可以定义先 y 后 x 的二次极限。

4. 二次极限和二重极限的关系: 二次极限存在时, 二重极限不一定存在; 反之, 二重极限存在时, 二次极限一定存在且等于二重极限。

5. 多元函数连续性: 若 $\lim_{\mathbf{x} \rightarrow \mathbf{x}_0} f(\mathbf{x}) = f(\mathbf{x}_0)$, 则称 $f(\mathbf{x})$ 在 \mathbf{x}_0 处连续; 若 $f(\mathbf{x})$ 在 D 上每一点均连续, 则称 f 是 D 上的连续函数。

6. 向量值函数: 设 D 是 \mathbb{R}^n 上的点集, D 到 \mathbb{R}^m 的映射 $f: D \rightarrow \mathbb{R}^m$ 称为 n 元 m 维向量值函数。

7. 连续函数的性质:

(1) 有界性定理: 若 K 为 \mathbb{R}^n 上的有界闭集, f 是 K 上的连续函数, 则

f 在 K 上有界;

(2) 最值定理: 设 K 是 \mathbb{R}^n 上的有界闭集, f 是 K 上的连续函数, 则 f 在 K 上必定能取到最大值和最小值;

(3) 一致连续: 若 K 是 \mathbb{R}^n 中的点集, $f: K \rightarrow \mathbb{R}^n$ 为映射. 若 $\forall \varepsilon > 0$, $\exists \delta > 0$, $\forall \mathbf{x}_1, \mathbf{x}_2 \in K$, $|\mathbf{x}_1 - \mathbf{x}_2| < \delta$, 均有 $|f(\mathbf{x}_1) - f(\mathbf{x}_2)| < \varepsilon$, 则称 $f(\mathbf{x})$ 在 K 上一致连续。

(4) 一致连续性定理: 若 K 是 \mathbb{R}^n 上的有界闭集, $f: K \rightarrow \mathbb{R}^n$ 为连续映射, 则 f 在 K 上一致连续。

1.7.3 偏导数与全微分

1. 定义: 设 $D \subset \mathbb{R}^2$ 为开集, $z = f(x, y)$ 为定义在 D 上的二元函数, $(x_0, y_0) \in D$ 为一定点。

(1) 关于 x 的偏导: 若极限

$$\lim_{\Delta x \rightarrow 0} \frac{f(x_0 + \Delta x, y_0) - f(x_0, y_0)}{\Delta x}$$

存在, 则称 f 在 (x_0, y_0) 点关于 x 可偏导, 极限值称为 f 关于 x 的偏导数。

(2) 关于 y 的偏导: 若极限

$$\lim_{\Delta y \rightarrow 0} \frac{f(x_0, y_0 + \Delta y) - f(x_0, y_0)}{\Delta y}$$

存在, 则称 f 在 (x_0, y_0) 点关于 y 可偏导, 极限值称为 f 关于 y 的偏导数。

(3) 偏导函数: 若 f 在任意 $(x_0, y_0) \in D$ 处关于 x 均可偏导, 则称 f 在 D 上可对 x 偏导, 得到的偏导函数记作 $\frac{\partial f}{\partial x}$ 或 f_x ; 同理, 可将 f 关于 y 的偏导记作 $\frac{\partial f}{\partial y}$ 或 f_y 。

2. **方向导数**: 设 $z = f(x, y)$ 为定义在 D 上的二元函数, $\boldsymbol{\nu} = (\cos \alpha, \sin \alpha)$ 为一个方向。记

$$\frac{\partial f}{\partial \boldsymbol{\nu}} = f_x \cos \alpha + f_y \sin \alpha$$

为 f 沿方向 $\boldsymbol{\nu}$ 的方向导数。

3. **全微分**: 若 z 的增量 Δz 可用 x 的增量、 y 的增量表示为

$$\Delta z = A\Delta x + B\Delta y + o(\sqrt{\Delta x^2 + \Delta y^2})$$

则称函数 $z = f(x, y)$ 是可微的, 并记线性主要部分 $A\Delta x + B\Delta y$ 为 $f(x, y)$ 的全微分, 记作 $dz = A dx + B dy$ 。其中 $A(x, y) = f_x$, $B(x, y) = f_y$ 。

4. **梯度**: 称向量 (f_x, f_y) 为函数 $z = f(x, y)$ 的梯度, 记作 $\mathbf{grad} f$ 。梯度表示了函数值增加最快的方向。

5. **高阶偏导数**: 对 f 求偏导数后再求偏导数的结果称为二阶偏导数, 对 f 求 n 次偏导的结果称为 n 阶偏导数。二阶偏导数按照对 x 和 y 求偏导的次序分为四种: f_{xx} 、 f_{xy} 、 f_{yx} 、 f_{yy} , 其中 $f_{xy} = f_{yx}$ 。

6. **雅可比行列式**: 考虑向量值函数

$$f(\mathbf{x}) = \begin{cases} y_1 = f_1(x_1, \cdots, x_n) \\ y_2 = f_2(x_1, \cdots, x_n) \\ \vdots \\ y_m = f_m(x_1, \cdots, x_n) \end{cases}$$

在每一个坐标分量 y_i 处, 对每一个自变量 x_j 求偏导数, 将结果用矩阵

$$J = \begin{pmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \vdots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{pmatrix}$$

表示, 该矩阵称为向量值函数的雅可比矩阵, 记作 $f'(\mathbf{x})$ 或 Df 。雅可比矩阵相当于向量值函数的导数。

7. 连续、可偏导和可微的关系: 可微必连续; 可微必可偏导; 可偏导不一定可微, 但偏导数连续时一定可微。

1.7.4 多元函数求导法则

1. 一般求导方法: 对 x 求偏导时, 把 y 视作常量, 仅对 x 求导; 对 y 求偏导时, 把 x 视作常量, 仅对 y 求导。

2. 多元复合函数的链式求导法则: 设 $z = z(x, y)$, 其中 $x = x(u, v)$, $y = y(u, v)$ 则有:

$$z_u = z_x \cdot x_u + z_y \cdot y_u z_v = z_x \cdot x_v + z_y \cdot y_v$$

面对更为复杂的情况时, 首先按照复合次序从上到下列出树状图, 再考虑自顶向下所有能连到目标变量的路径, 求处此路径上所有偏导数的乘积, 最终将它们求和。

3. 一阶全微分的形式不变性: 无论 x 、 y 是自变量还是中间变量, 一阶全微分 $dz = z_x dx + z_y dy$ 这一形式始终不变。

4. 隐函数存在性定理: 若二元函数 $F(x, y)$ 满足 $F(x_0, y_0) = 0$, 在以 (x_0, y_0)

为中心的某一闭矩形上连续且具有连续偏导数, 同时 $F_y(x_0, y_0) \neq 0$, 则可以从隐函数 $F(x, y) = 0$ 中唯一确定隐函数 $y = f(x)$ 。

5. 多元隐函数求导方法: 对 $F(y, x_1, x_2, \dots, x_n)$ 两边分别关于 x_1, x_2, \dots, x_n 求导, 将 $\frac{\partial y}{\partial x_1}, \dots, \frac{\partial y}{\partial x_n}$ 视作变量, 求解 n 元方程组。

1.7.5 多元函数微分学的应用

1. 中值定理: 设 $f(x, y)$ 在凸区域 $D \subset \mathbb{R}^2$ 上可微, 则对 D 上任意两点 (x_0, y_0) 和 $x_0 + \Delta x, y_0 + \Delta y$, 存在 $0 < \theta < 1$, 记 $\mathbf{t} = (x_0 + \theta\Delta x, y_0 + \theta\Delta y)$, 则

$$f(x_0 + \Delta x, y_0 + \Delta y) - f(x_0, y_0) = f_x(\mathbf{t})\Delta x + f_y(\mathbf{t})\Delta y$$

2. 泰勒展开: 设 $f(x, y)$ 在点 (x_0, y_0) 的邻域 U 上具有 $n + 1$ 阶连续偏导数, 则 U 内每一点成立

$$f(x_0 + \Delta x, y_0 + \Delta y) = \sum_{k=0}^n \frac{1}{k!} \left(\frac{\partial}{\partial x} \Delta x + \frac{\partial}{\partial y} \Delta y \right)^k f(x_0, y_0) + o\left((\Delta x^2 + \Delta y^2)^{\frac{k}{2}}\right)$$

3. 空间曲线的切线和法平面:

(1) 空间曲线的参数方程: 设参数 $t \in [a, b]$, 则参数方程

$$\Gamma : \begin{cases} x = x(t) \\ y = y(t) \\ z = z(t) \end{cases}$$

表示空间中的一条曲线。

(2) 空间曲线的切向量: 向量 $\mathbf{r}'(t_0) = (x'(t_0), y'(t_0), z'(t_0))$ 是曲线 Γ 在 $t = t_0$

处的切向量。

(3) 空间曲线的切线方程：当 $t = t_0$ 时，空间曲线 Γ 的切线方程为

$$\frac{x - x_0}{x'(t_0)} = \frac{y - y_0}{y'(t_0)} = \frac{z - z_0}{z'(t_0)}$$

(4) 空间曲线的法平面：过定点 $(x(t_0), y(t_0), z(t_0))$ 且与该点切线垂直的平面，方程为

$$x'(t_0)(x - x_0) + y'(t_0)(y - y_0) + z'(t_0)(z - z_0) = 0$$

4. 曲面的切平面和法线：

(1) 空间曲面方程： $S : F(x, y, z) = 0$ 。

(2) 空间曲面的切平面：在 $P_0(x_0, y_0, z_0)$ 处，空间曲面 S 的切平面方程为

$$F_x(P_0)(x - x_0) + F_y(P_0)(y - y_0) + F_z(P_0)(z - z_0) = 0$$

(3) 空间曲面的法向量： $\mathbf{n} = (F_x(P_0), F_y(P_0), F_z(P_0))$ 。

(4) 空间曲面的法线：

$$\frac{x - x_0}{F_x(P_0)} = \frac{y - y_0}{F_y(P_0)} = \frac{z - z_0}{F_z(P_0)}$$

6. 求函数最值：设 nabla 算子 $\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}\right)$ ，则求函数 $f(x, y)$ 的最值时，先求解 $\nabla f = \mathbf{0}$ 得到 f 的所有极值，再比较得出 f 的最大值和最小值。

7. 条件极值：设函数 $f = f(x_1, x_2, \dots, x_n)$ 满足约束 $G_i(x_1, \dots, x_n) = 0$ ，其中 $i = 1, 2, \dots, m$ ，构建 Lagrange 函数

$$L(x_1, \dots, x_n, \lambda_1, \dots, \lambda_m) = f(x_1, \dots, x_n) + \sum_{i=1}^m \lambda_i G_i(x_1, \dots, x_n)$$

列出 $m+n$ 个方程: $\frac{\partial L}{\partial x_i} = 0$ 与 $\frac{\partial L}{\partial \lambda_i} = 0$, 从中解出极值点 (x_1, \dots, x_n) 满足的条件。此方法称为 Lagrange 乘数法。

1.8 多元函数积分学

和之前的一元函数微积分略有区别, 学习多元函数积分学的重心不在于理论证明, 而更注重实际应用。多元函数积分对现实问题的刻画更为深刻, 每一种线面积分都有实际的物理意义, 在物理、工程上有非常广泛的应用。

当积分区域从平面变为空间时, 首先要解决的是最简单的问题, 即在 xOy 平面的投影是矩形的情况, 这就引入了二重积分。若为空间里的每个点赋予密度, 对空间几何体求质量的问题就可转化为三重积分。因此, 学习多元函数积分学的第一步, 是掌握重积分的各种技巧。

若积分对象变得不规则, 如空间中的任意曲线、曲面, 此时需要引入另一类积分, 那就是线面积分。若积分时不区分方向, 则称为对弧长/面积的线面积分, 即第一类线面积分; 若积分时考虑方向, 则称为对坐标的线面积分, 即第二类线面积分。作为物理中的重要应用, 还介绍了部分关于场论的概念。

在数学分析中有一个非常优美的公式——流形上的 Stokes 公式, 它统一了微积分基本定理、格林公式、高斯公式和斯托克斯公式。为了介绍该著名公式, 额外引入了微分形式与外微分作为铺垫。

最后介绍了含参变量积分。若对二元函数中的某一个变量积分, 得到的便是关于另一个变量的函数, 这种函数被称为含参变量积分。它是数学分析中某些后续课程的基础, 因此有必要介绍含参变量积分的相关性质。

1.8.1 重积分

1. **二重积分**: 考虑一个曲顶柱体, 底面是 xOy 平面上的有界闭区域 D , 顶面是非负连续函数 $z = f(x, y)$, 则该曲顶柱体的体积便是 $f(x, y)$ 在区域 D 上的二重积分, 记作

$$\iint_D f(x, y) d\sigma$$

其中 $d\sigma$ 为面积微元。

2. 二重积分的计算:

(1) 直角坐标系上: $d\sigma$ 可用 $dx dy$ 表示。

(2) 先 y 后 x 积分法: 若 D 的横坐标取值范围是 $[a, b]$, 对任意 $x_0 \in [a, b]$, D 与 $x = x_0$ 所交线段纵坐标在 $[f(x_0), g(x_0)]$ 内, 则

$$\iint_D f(x, y) dx dy = \int_a^b dx \left[\int_{f(x)}^{g(x)} f(x, y) dy \right]$$

(3) 先 x 后 y 积分法: 若 D 的纵坐标取值范围是 $[c, d]$, 对任意 $y_0 \in [c, d]$, D 与 $y = y_0$ 所交线段横坐标在 $[f(y_0), g(y_0)]$ 内, 则

$$\iint_D f(x, y) dx dy = \int_c^d dy \left[\int_{f(y)}^{g(y)} f(x, y) dx \right]$$

3. 二重积分的变量代换:

(1) 极坐标代换公式: 设 $x = r \cos \theta$, $y = r \sin \theta$, 其中 $\theta \in [\theta_1, \theta_2]$, r 的取值范围为 $[r_1(\theta), r_2(\theta)]$, 则

$$\iint_D f(x, y) dx dy = \int_{\theta_1}^{\theta_2} d\theta \left[\int_{r_1(\theta)}^{r_2(\theta)} f(x, y) \cdot r dr \right]$$

(2) 一般坐标代换公式: 设 $x = x(u, v)$, $y = y(u, v)$, 雅可比行列式

$$\frac{\partial(x, y)}{\partial(u, v)} = \begin{vmatrix} x_u & x_v \\ y_u & y_v \end{vmatrix}$$

则面积微元可按照如下公式代换:

$$dxdy = \left| \frac{\partial(x, y)}{\partial(u, v)} \right| du dv$$

4. 三重积分: 考虑空间上的封闭几何体 Ω , 每一点 (x, y, z) 处的密度为 $f(x, y, z)$, 则该几何体的质量便是 $f(x, y, z)$ 在空间几何体 Ω 上的三重积分, 记作

$$\iiint_{\Omega} f(x, y, z) dV$$

5. 三重积分的计算:

(1) 直角坐标系上: dV 可用 $dxdydz$ 表示。

(2) 先 1 后 2 积分法: 先求空间几何体在平面上的投影 D (以 xOy 平面为例), 过 D 内任一点 (x, y) 作 z 轴平行线, 与空间几何体所交线段的 z 坐标范围为 $[z_1, z_2]$, 则

$$\iiint_{\Omega} f(x, y, z) dxdydz = \iint_D dxdy \left[\int_{z_1}^{z_2} f(x, y, z) dz \right]$$

(3) 先 2 后 1 积分法: 设空间几何体的 z 坐标范围为 $[z_1, z_2]$, 过任意

$z \in [z_1, z_2]$ 作平行于 xOy 的平面, 交空间几何体于平面区域 $D(z)$, 则

$$\iiint_{\Omega} f(x, y, z) dx dy dz = \int_{z_1}^{z_2} dz \left[\iint_{D(z)} f(x, y, z) dx dy \right]$$

6. 三重积分的变量代换:

(1) 柱坐标代换公式: 令 $x = r \cos \theta$, $y = r \sin \theta$, z 保持不变, 则

$$\iiint_{\Omega} f(x, y, z) dx dy dz = \iiint_{\Omega} f(r, \theta, z) r dr d\theta dz$$

(2) 球坐标代换公式: 令 $x = r \sin \varphi \cos \theta$, $y = r \sin \varphi \sin \theta$, $z = r \cos \varphi$, 则

$$\iiint_{\Omega} f(x, y, z) dx dy dz = \iiint_{\Omega} f(r, \theta, \varphi) r^2 \sin \varphi dr d\theta d\varphi$$

其中 φ 表示和 z 轴负半轴的夹角范围, 通常取 $[0, \pi]$; θ 表示 xOy 平面上旋转角度, 通常取 $[0, 2\pi]$ 。

(3) 一般坐标代换公式: 令 x, y, z 分别为 u, v, w 的函数, 则

$$\iiint_{\Omega} f(x, y, z) dx dy dz = \iiint_{\Omega} f(u, v, w) \left| \frac{\partial(x, y, z)}{\partial(u, v, w)} \right| du dv dw$$

1.8.2 曲线积分

1. **第一类曲线积分：**对弧长的曲线积分。考虑一段各点密度 $f(x, y)$ 已知的平面曲线 L ，对其求质量时，取弧长微元 ds ，则有第一类曲线积分：

$$\int_L f(x, y) ds$$

当 L 为空间曲线时，弧长微元仍为 ds ，第一类曲线积分定义为 $\int_L f(x, y, z) ds$ 。

2. **第一类曲线积分求法：**平面曲线有 $ds = \sqrt{(x')^2 + (y')^2} dx dy$ ；空间曲线有 $ds = \sqrt{(x')^2 + (y')^2 + (z')^2} dx dy dz$ 。

3. **第二类曲线积分：**对坐标的曲线积分。考虑一段标定方向的空间曲线 L ，以力 $f(x, y) = (P(x, y), Q(x, y))$ 沿着 L 做功，则有第二类曲线积分：

$$\int_L P(x, y) dx + Q(x, y) dy$$

空间曲线上的第二类曲线积分定义为 $\int_L P(x, y, z) dx + Q(x, y, z) dy + R(x, y, z) dz$ 。

4. **第二类曲线积分求法：**用参数 t 描述曲线 L ，则可以代入 $dx = x'(t) dt$ ，同理有 $dy = y'(t) dt$ 和 $dz = z'(t) dt$ 。

5. **格林公式：**设 D 为平面上光滑或分段光滑的简单闭曲线所围成的单连通闭区域， ∂D 为逆时针方向的区域边界，且函数 $P(x, y)$ 和 $Q(x, y)$ 在 D 上具有连续偏导数，则有

$$\oint_{\partial D} P dx + Q dy = \iint_D \left(\frac{\partial Q}{\partial x} - \frac{\partial P}{\partial y} \right) dx dy$$

6. **曲线积分与路径无关的条件：**若 $P dx + Q dy$ 恰好为函数 $u(x, y)$ 的全微分，

即

$$du = Pdx + Qdy$$

则曲线积分 $\int_L Pdx + Qdy$ 与路径 L 无关, 仅与起点 A 和终点 B 有关, 且积分值为 $u(x, y)|_A^B$ 。判定方式: $\frac{\partial P}{\partial y} = \frac{\partial Q}{\partial x}$ 。

1.8.3 曲面积分

1. 第一类曲面积分: 对面积的曲面积分。考虑各点密度 $f(x, y, z)$ 已知的空间曲面 Σ , 对其求质量时, 取面积微元 dS , 则有第一类曲面积分:

$$\iint_{\Omega} f(x, y, z) dS$$

2. 第一类曲面积分求法: 当曲面方程可用 $z = z(x, y)$, $(x, y) \in D$ 表示时, 面积微元 $dS = \sqrt{1 + (z_x)^2 + (z_y)^2} dx dy$, 即

$$\iint_{\Omega} f(x, y, z) dS = \iint_D f(x, y, z(x, y)) \sqrt{1 + (z_x)^2 + (z_y)^2} dx dy$$

3. 第二类曲面积分: 对坐标的曲面积分。考虑各点流速 (P, Q, R) 已知的流体在定向曲面 Σ 上的流量, 则有第二类曲面积分:

$$\iint_{\Omega} P(x, y, z) dy dz + Q(x, y, z) dz dx + R(x, y, z) dx dy$$

4. 第二类曲面积分求法: 首先确定 Ω 定向。以 $P dx dy$ 分量为例, 用平行于 z 轴正方向的线穿过曲面, 若该线穿入曲面则定为反向 (二重积分变号), 穿出曲面则定为正向。定向后, 分别将 $P dx dy$ 、 $Q dy dz$ 、 $R dz dx$ 投影到 xOy 、 yOz 、 zOx

平面, 分别在各投影面上直接用二重积分的方法求解各个分量的积分值, 最后求和。此方法麻烦且不常用, 一般都可以用后面介绍的高斯公式求解。

5. **高斯公式:** 设 Ω 是由光滑或分片光滑的封闭曲面围成的闭区域, 函数 $P(x, y, z)$, $Q(x, y, z)$, $R(x, y, z)$ 在 Ω 上具有连续偏导数, $\partial\Omega$ 为 Ω 的外侧曲面, 则

$$\oiint_{\partial\Omega} Pdydz + Qdzdx + Rdxdy = \iiint_{\Omega} \left(\frac{\partial P}{\partial x} + \frac{\partial Q}{\partial y} + \frac{\partial R}{\partial z} \right) dxdydz$$

6. **斯托克斯公式:** 设 Σ 为光滑曲面, 其边界 $\partial\Sigma$ 为分段光滑闭曲线。若函数 $P(x, y, z)$, $Q(x, y, z)$, $R(x, y, z)$ 在 Σ 以及 $\partial\Sigma$ 上具有连续偏导数, 则成立

$$\int_{\partial\Sigma} Pdx + Qdy + Rdz = \iint_{\Sigma} \begin{vmatrix} dydz & dzdx & dxdy \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}$$

1.8.4 流形上的 Stokes 公式

1. **向量外积:** 设 $\mathbf{a} = (a_1, a_2)$, $\mathbf{b} = (b_1, b_2)$, 则定义向量外积

$$\mathbf{a} \wedge \mathbf{b} = \begin{vmatrix} a_1 & a_2 \\ b_1 & b_2 \end{vmatrix}$$

外积具有反对称性和对加法的分配律, 且 $\mathbf{a} \wedge \mathbf{a} = 0$ 。向量外积的几何意义为平行四边形的有向面积。

2. **微分形式:** 以微分及其外积作为一组基的向量空间。

(1) 一次微分形式: 令 $\mathbf{x} = (x_1, \cdots, x_n)$, 则 1- 形式为

$$a_1(\mathbf{x})dx_1 + \cdots + a_n(\mathbf{x})dx_n \in \Lambda^1$$

(2) 二次微分形式: 由反对称性得 $dx_i \wedge dx_j = -dx_j \wedge dx_i$, 则 2- 形式为

$$\sum_{1 \leq i < j \leq n} g_{ij}(\mathbf{x}) dx_i \wedge dx_j \in \Lambda^2$$

3. 微分形式的外积: 令 $\omega = a_1(\mathbf{x})dx_1 + \cdots + a_n(\mathbf{x})dx_n$, $\eta = b_1(\mathbf{x})dx_1 + \cdots + b_n(\mathbf{x})dx_n$:

$$\begin{aligned} \omega \wedge \eta &= \sum_{i,j=1}^n a_i(\mathbf{x})b_j(\mathbf{x})dx_i \wedge dx_j \\ &= \sum_{1 \leq i < j \leq n} \begin{vmatrix} a_i(\mathbf{x}) & a_j(\mathbf{x}) \\ b_i(\mathbf{x}) & b_j(\mathbf{x}) \end{vmatrix} dx_i \wedge dx_j \end{aligned}$$

4. 外微分: 对 Λ^1 的任意 1- 形式 $\omega = a_1(\mathbf{x})dx_1 + \cdots + a_n(\mathbf{x})dx_n$, 定义 $d\omega = da_1(\mathbf{x}) \wedge dx_1 + \cdots + da_n(\mathbf{x}) \wedge dx_n$ 。同理, 当 ω 为 2- 形式时, 定义

$$d\omega = \sum_{1 \leq i < j \leq n} dg_{ij}(\mathbf{x}) \wedge dx_i \wedge dx_j$$

5. 流形上的斯托克斯公式: 高次微分形式 $d\omega$ 在给定区域上的积分等于低一次的微分形式 ω 在低一维的区域边界上的积分。定义 M 为微分流形, ∂M 为 M 具有诱导定向的边界, 则

$$\int_{\partial M} \omega = \int_M d\omega$$

(1) Newton-Leibniz 公式: 令 $M = [a, b]$, 则 $\partial M = \big|_a^b$ 。此时

$$F(x)\big|_a^b = \int_a^b dF(x) = \int_a^b f(x)dx$$

(2) Green 公式: 令 M 为平面封闭区域 D , 取诱导定向为逆时针, 此时

$$\begin{aligned}\oint_{\partial D} Pdx + Qdy &= \iint_D (P_x dx + P_y dy) \wedge dx + (Q_x dx + Q_y dy) \wedge dy \\ &= \iint_D (Q_y - P_x) dx \wedge dy\end{aligned}$$

(3) Gauss 公式: 令 M 为空间封闭区域 Ω , 取诱导定向为外侧, 此时

$$\begin{aligned}\oiint_{\partial\Omega} Pdy \wedge dz + Qdz \wedge dx + Rdx \wedge dy \\ &= \iiint_{\Omega} (P_x dx) \wedge dy \wedge dz + (Q_y dy) \wedge dz \wedge dx + (R_z dz) \wedge dx \wedge dy \\ &= \iiint_{\Omega} (P_x + Q_y + R_z) dx \wedge dy \wedge dz\end{aligned}$$

(4) Stokes 公式: 令 M 为空间曲面 S , 取诱导定向为逆时针, 此时

$$\begin{aligned}\oint_{\partial D} Pdx + Qdy + Rdz &= \iint_D (P_y \wedge dy + P_z \wedge dz) \wedge dx \\ &\quad + (Q_x \wedge dx + Q_z \wedge dz) \wedge dy + (Q_x \wedge dx + R_y \wedge dy) \wedge dz \\ &= \iint_D \begin{vmatrix} dy \wedge dz & dz \wedge dx & dx \wedge dy \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P & Q & R \end{vmatrix}\end{aligned}$$

1.8.5 场论

1. **场**: 设 $\Omega \subset \mathbb{R}^3$ 为一个区域, 若在 t 时刻, Ω 中每一点 (x, y, z) 都有一个确定的 $f(x, y, z, t)$ 与之对应, 则称 f 为 Ω 上的场。当 f 为多元函数时, 称为数量场; 当 f 为向量值函数时, 称为向量场。

2. **梯度**: 若 $f(x, y, z)$ 在 Ω 上具有连续偏导数, 则定义 f 的梯度为

$$\mathbf{grad} f = f_x \mathbf{i} + f_y \mathbf{j} + f_z \mathbf{k}$$

函数沿梯度方向上升最快。

3. **散度**: 设 $\mathbf{a} = (P, Q, R)$ 为 Ω 上的向量场, M 为 Ω 中一点, 定义

$$\operatorname{div} \mathbf{a}(M) = P_x(M) + Q_y(M) + R_z(M)$$

为向量场 \mathbf{a} 在 M 点的散度。当散度为正时, M 为源点; 当散度为负时, M 为汇点; 若 Ω 中任一点的散度均为 0, 则称 Ω 为无源场。

4. **旋度**: 设 $\mathbf{a} = (P, Q, R)$ 为 Ω 上的向量场, M 为 Ω 中一点, 定义

$$\operatorname{rot} \mathbf{a}(M) = \begin{vmatrix} \mathbf{i} & \mathbf{j} & \mathbf{k} \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ P(M) & Q(M) & R(M) \end{vmatrix}$$

为向量场 \mathbf{a} 在 M 点的旋度。当旋度处处为 0 时, \mathbf{a} 称为保守场, 此时 \mathbf{a} 中的曲线积分与路径无关。

5. Hamilton 算子:

(1) Nabla 算子: 定义如下, 且 $\nabla f = \mathbf{grad} f$, $\nabla \cdot \mathbf{a} = \text{div } \mathbf{a}$, $\nabla \times \mathbf{a} = \mathbf{rot } \mathbf{a}$ 。

$$\nabla = \mathbf{i} \frac{\partial}{\partial x} + \mathbf{j} \frac{\partial}{\partial y} + \mathbf{k} \frac{\partial}{\partial z}$$

(2) Laplace 算子: 定义如下, 且 $\Delta = \nabla \cdot \nabla$, 称 $\Delta u = 0$ 为调和方程。

$$\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2}$$

1.8.6 含参变量积分

1. 定义: 设 $f(x)$ 是定义在闭矩形 $[a, b] \times [c, d]$ 上的连续函数, 则定义关于 y 的函数

$$I(y) = \int_a^b f(x, y) dx$$

定义域为 $[c, d]$ 。同样可以定义关于 x 的函数

$$J(x) = \int_c^d f(x, y) dy$$

定义域 $[a, b]$ 。这种对 $f(x, y)$ 中某一个变量积分得到的函数称为含参变量积分。

2. 含参变量积分的分析性质: 若 $f(x)$ 在闭矩形上是连续函数, 则积分得到的函数连续, 积分号可分别与求导、极限、积分交换次序。

3. 含参变量积分求导: 设 $F(y) = \int_{a(y)}^{b(y)} f(x, y) dx$, 则

$$F'(y) = \int_{a(y)}^{b(y)} f_y(x, y) dx + b'(y)f(b, y) - a'(y)f(a, y)$$

4. 含参变量广义积分:

(1) 一致收敛定义: 设 $f(x, y)$ 的定义域为 $[a, +\infty) \times [c, d]$ 。若 $\forall \varepsilon > 0$,

$\exists A_0 > 0, \forall A > A_0, \forall y \in [c, d],$

$$\left| \int_a^A f(x, y) dx - I(y) \right| < \varepsilon$$

则称 $\int_a^{+\infty} f(x, y) dx$ 关于 y 在 $[c, d]$ 上一致收敛于 $I(y)$ 。

(2) 柯西收敛准则: $\int_a^{+\infty} f(x, y) dx$ 关于 y 在 $[c, d]$ 上一致收敛的充要条件:
 $\forall \varepsilon > 0, \exists A_0 > 0, \forall A_2 > A_1 > A_0, \forall y \in [c, d],$

$$\left| \int_{A_1}^{A_2} f(x, y) dx \right| < \varepsilon$$

(3) Weierstrass 判别法: 若存在 $F(x)$ 使得 $|f(x, y)| \leq F(x)$ 在定义域内恒成立, 且 $\int_a^{+\infty} F(x) dx$ 收敛, 则 $\int_a^{+\infty} f(x, y) dx$ 在 $[c, d]$ 上一致收敛。

(4) A-D 判别法: 若下面某条件满足, 则含参变量积分 $\int_a^{+\infty} f(x, y) g(x, y) dx$ 在 $[c, d]$ 上一致收敛:

· Abel 条件: $\int_a^{+\infty} f(x, y) dx$ 在 $[c, d]$ 上一致收敛, 且对任意 $y \in [c, d], g(x, y)$ 关于 x 单调且一致有界;

· Dirichlet 条件: $\int_a^A f(x, y) dx$ 在 $[c, d]$ 上关于 A 一致有界, 且对任意 $y \in [c, d], g(x, y)$ 关于 x 单调且一致趋于 0。

第二章 高等代数

Advanced Algebra

高等代数同为大一新生必修的数院基础课,与数分一样,是多门后继学科最基本的理论基础. 山大的高代分为两个学期讲授: 第一个学期主要学习线性代数基础, 第二个学期主要学习线性变换和矩阵分析. 高代 2 被称为挂科神课不无道理, 相比前一学期内容跨度相当大, 瞬间抽象起来, 引入的线性空间和线性映射等抽象概念很难理解. 但是学得深入又会感觉非常有趣, 只要满足八大性质的集合, 就能用一组基和常见的欧氏空间和坐标联系起来, 充分体现了数学的抽象之美.

学习高等代数, 不需要任何先修知识, 但是要学会将复杂的事物抽象化描述 (最好的例子: 线性方程组改写为 $Ax = b$). 初学阶段, 多思考, 多举例子, 多做题对提高很有帮助.

高代 1 的重点: 多项式, 行列式, 矩阵, 线性方程组, 二次型

高代 2 的重点: 线性空间, 线性映射与线性变换, 特征值与特征向量, λ 矩阵与 Jordan 标准形, 欧氏空间, 正交变换

2.1 多项式

2.1.1 数域与一元多项式

1. 设 P 是复数集的一个子集, 且 P 对四则运算封闭, 则称 P 为一个数域。
2. 一元多项式: 设 n 为非负整数, $a_1, \dots, a_n \in P$, 则称

$$f(x) = a_n x^n + a_{n-1} x^{n-1} + \dots + a_1 x + a_0$$

为数域 P 上的一元多项式。

3. 多项式的相等: 若 $f(x) = a_n x^n + \dots + a_1 x + a_0$, $g(x) = b_n x^n + \dots + b_1 x + b_0$, 且 $a_i = b_i$, 则称 $f(x)$ 与 $g(x)$ 相等, 记作 $f(x) = g(x)$ 。
4. 多项式的次数: 若 $a_n \neq 0$, 则称 $f(x)$ 为 n 次多项式, 记作 $\partial(f) = n$, 且称 a_n 为 $f(x)$ 的首项系数。当 $a_n = 1$ 时, 称 $f(x)$ 为首一多项式。
5. 一元多项式环: 所有数域 P 上的一元多项式全体, 记作 $P[x]$ 。

2.1.2 一元多项式的因式

1. 多项式除法: 若 $\partial(f) \geq \partial(g)$, 且存在 $q(x), r(x) \in P[x]$, 使得 $f(x) = q(x) \cdot g(x) + r(x)$, 满足 $\partial(r) < \partial(g)$, 则称 $q(x)$ 为 $g(x)$ 除 $f(x)$ 的商, $r(x)$ 为 $g(x)$ 除 $f(x)$ 的余。对固定的 $f(x)$ 和 $g(x)$, 商和余均是唯一的。
2. 整除: 若 $r(x) = 0$, 则称 $g(x)$ 整除 $f(x)$, 记作 $g(x) \mid f(x)$ 。此时称 $g(x)$ 为 $f(x)$ 的因式。
3. 公因式: 若 $\exists h(x) \in P[x]$, 使得 $h(x) \mid f(x)$ 且 $h(x) \mid g(x)$, 则称 $h(x)$ 为 $f(x)$ 和 $g(x)$ 的公因式。
4. 最大公因式: 若 $h(x)$ 为 $f(x)$ 和 $g(x)$ 的公因式, 且对任意 $f(x)$ 和 $g(x)$ 的公因式 $h_1(x)$, 均有 $h_1(x) \mid h(x)$, 则称 $h(x)$ 为 $f(x)$ 的最大公因式, 记作 $h = (f, g)$ 。

5. 最大公因式表示定理：若 $h = (f, g)$ ，则 $\exists u(x), v(x) \in P[x]$ ，使 $h(x) = u(x) \cdot f(x) + v(x) \cdot g(x)$ 。

6. Euclid 辗转相除法：设 $f(x), g(x) \in P[x]$ ，以如下步骤求最大公因式：

(1) 若 $\partial(f) = \partial(g)$ ，则通过数乘和减法使得 $\partial(f) > \partial(g)$ ；

(2) 用商和余表示： $f(x) = q(x) \cdot g(x) + r(x)$ ；

(3) 若 $r(x) = 0$ ，则 $g(x)$ 即为最大公因式，退出程序；

(4) 否则，用 $g(x)$ 代替 $f(x)$ ， $r(x)$ 代替 $g(x)$ ，返回第 (2) 步。

7. 互素多项式：若 $(f, g) = 1$ ，则称 $f(x)$ 和 $g(x)$ 为互素多项式。

2.1.3 多项式的因式分解

1. 不可约多项式：因式只有 1 和本身的多项式。

2. 因式分解定理： $\forall f(x) \in P[x]$ ，且 $\partial(f) > 1$ ，则存在唯一分解式

$$f(x) = c \cdot p_1(x) \cdot p_2(x) \cdots p_s(x)$$

其中 $p_i(x)$ 均为首一不可约多项式， $c \in P$ 为常数。

3. 重因式：若 $r^k(x) \mid f(x)$ ，而 $r^{k+1}(x) \nmid f(x)$ ，则称 $r(x)$ 为 $f(x)$ 的 k 重因式。

4. 代数基本定理：对每个次数 ≥ 1 的复系数多项式，在复数域上一定有一个一次因式。

2.1.4 有理系数多项式

1. 定义： $\mathbb{Q}[x]$ 上的多项式。

2. 性质：每个次数 ≥ 1 的有理系数多项式都能唯一分解为不可约有理系数多项式的乘积。

3. **本原多项式**: 在 $g(x) = b_n x^n + \cdots + b_1 x_1 + b_0$ 中, 若 b_n, \cdots, b_0 互素, 则称 $g(x)$ 为本原多项式。

4. **求整系数多项式全部有理根**: 若 $f(x) \in \mathbb{Z}[x]$, 而 $\frac{r}{s}$ 为一个有理根 (其中 r, s 互素), 则必有 $r \mid a_0, s \mid a_n$, 枚举全部组合再验证即可。

5. **Eisenstein 判别法**: 设 $f(x) \in \mathbb{Z}[x]$, 若存在素数 p , 使得

$$(1) p \nmid a_n; (2) p \nmid a_i, i = 1, 2, \cdots, n-1; (3) p^2 \nmid a_0,$$

则 $f(x)$ 在有理数上不可约。

2.2 向量与线性方程组

2.2.1 线性方程组

1. **定义**: 形如

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + \cdots + a_{2n}x_n = b_2 \\ \vdots \quad \quad \quad \vdots \quad \quad \quad \vdots \\ a_{m1}x_1 + \cdots + a_{mn}x_n = b_m \end{cases}$$

的方程组称为 n 元线性方程组。

2. **齐次线性方程组**: 当常数项 $b_1 = \cdots = b_n = 0$ 时, 称上述方程组为

3. **线性方程组的解**: 若存在一组实数 (x_1, \cdots, x_n) , 使得全部 m 个等式成立, 则称 (x_1, \cdots, x_n) 为线性方程组的一组解。线性方程组全部解的集合称为解集。

4. **初等变换**: 有下列三种形式:

- (1) 用某一非零的数乘以该方程;
- (2) 把一个方程的倍数加到另一个方程;
- (3) 互换两个方程的位置。

初等变换不改变方程组的解集。

5. 高斯消元法：将线性方程组化为阶梯型（下式中令 $m > n$ ）：

$$\left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ \quad a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \qquad \qquad \qquad \vdots \qquad \qquad \vdots \\ \qquad \qquad \qquad a_{mn}x_n = b_n \\ \qquad \qquad \qquad 0 = 0 \\ \qquad \qquad \qquad \qquad \qquad \vdots \\ \qquad \qquad \qquad 0 = 0 \end{array} \right.$$

自下而上依次求解 $x_n, x_{n-1}, \cdots, x_1$ 。

6. 线性方程组解的个数：称“ $0 = 0$ ”以外的方程为有效方程，个数为 r 。

（1）无解：若存在方程“ $0 = c$ ”，其中常数 $c \neq 0$ ，则方程组无解；

（2）恰有一组解：若 $r = n$ ，则方程组恰有一组解；

（3）无穷多组解：若 $r < n$ ，则方程组有无穷多组解，每组解均能用 $n - r$ 个自由变量表示。

2.2.2 向量空间

1. 向量： \mathbb{R}^n 中的元素 $\mathbf{x} = (x_1, \cdots, x_n)$ 称为向量，可用于表示 n 元线性方程组的一组解。

2. n 维向量空间：在 \mathbb{R}^n 上定义加法和数乘构成的空间。令 $\mathbf{x} = (x_1, \cdots, x_n)$, $\mathbf{y} = (y_1, \cdots, y_n)$ ，则

（1）加法： $\mathbf{x} + \mathbf{y} = (x_1 + y_1, \cdots, x_n + y_n)$ ；

(2) 数乘: $\lambda x = (\lambda x_1, \dots, \lambda x_n)$;

(3) 点积: $\mathbf{x} \cdot \mathbf{y} = x_1 y_1 + x_2 y_2 + \dots + x_n y_n$ 。

3. 线性组合: 若 $\exists k_1, \dots, k_n \in \mathbb{R}$, 使 $\boldsymbol{\beta} = k_1 \boldsymbol{\alpha}_1 + \dots + k_n \boldsymbol{\alpha}_n$, 则称 $\boldsymbol{\beta}$ 为 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n$ 的一个线性组合, 也称 $\boldsymbol{\beta}$ 可被 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n$ 线性表示。

4. 线性相关: 若存在不全为零的 $k_1, \dots, k_n \in \mathbb{R}$, 使得

$$k_1 \boldsymbol{\alpha}_1 + \dots + k_n \boldsymbol{\alpha}_n = \mathbf{0}$$

则称 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n$ 线性相关。

5. 线性无关: 若 $k_1 \boldsymbol{\alpha}_1 + \dots + k_n \boldsymbol{\alpha}_n = \mathbf{0} \Rightarrow k_1 = k_2 = \dots = k_n = 0$, 则称 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n$ 线性无关。

6. 极大线性无关组: 若 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n$ 线性无关, 且 $\forall k = n+1, \dots, m$, 均有 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n, \boldsymbol{\alpha}_k$ 线性相关, 则称 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n$ 为向量组 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_n, \boldsymbol{\alpha}_{n+1}, \dots, \boldsymbol{\alpha}_m$ 的一个极大线性无关组, 其中的向量个数称为 $\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m$ 的秩, 记作 $\text{rank}(\boldsymbol{\alpha}_1, \dots, \boldsymbol{\alpha}_m) = n$ 。

7. 重要结论*: n 维向量组的秩不超过 n , 即 n 个线性无关的 n 维向量必定可以线性表示所有 n 维向量。

2.2.3 线性方程组解的结构

1. 解的性质: 同一方程组的解经过加法、数乘运算, 仍是原方程组的解。

2. 基础解系: 一组线性无关的解向量 $\boldsymbol{\eta}_1, \dots, \boldsymbol{\eta}_t$, 且能线性表示方程组的所有解。

3. 导出组: 将 b_1, \dots, b_n 换为 0, 得到对应的齐次线性方程组称为原方程组的导出组。

4. 非齐次线性方程组的通解: 设 $\boldsymbol{\gamma}_0$ 为原方程组的一组解, 则所有解可表示

为 $\boldsymbol{\gamma} = \boldsymbol{\gamma}_0 + \boldsymbol{\eta}$, 其中 $\boldsymbol{\eta}$ 是导出组的任意解。

2.3 矩阵与行列式

2.3.1 矩阵的概念

1. 定义：由 m 个 n 维行向量组成的数表称为 $m \times n$ 矩阵，记作

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}$$

为了表述方便，通常还会使用如下两种表述矩阵的方式：

(1) 元素、下标表示法： $\mathbf{A} = (a_{ij})_{m \times n}$ 。

(2) 列向量表示法：令 \mathbf{A} 的 n 个列向量分别为 $\boldsymbol{\alpha}_j = (a_{1j}, a_{2j}, \cdots, a_{mj})^T$ ，则矩阵 \mathbf{A} 可表示为 $\mathbf{A} = (\boldsymbol{\alpha}_1, \cdots, \boldsymbol{\alpha}_n)$ 。

2. 矩阵的加法：令 $\mathbf{A} = (a_{ij})_{m \times n}$, $\mathbf{B} = (b_{ij})_{m \times n}$, $\mathbf{C} = \mathbf{A} + \mathbf{B}$, 则 $c_{ij} = a_{ij} + b_{ij}$ 。

3. 矩阵的数乘：令 $\lambda \in \mathbb{R}$, $\mathbf{A} = (a_{ij})_{m \times n}$, $\mathbf{B} = \lambda \cdot \mathbf{A}$, 则 $b_{ij} = \lambda \cdot a_{ij}$ 。

4. 矩阵乘法：令 $\mathbf{A} = (a_{ij})_{m \times n}$, $\mathbf{B} = (b_{ij})_{n \times s}$, $\mathbf{C} = \mathbf{AB}$, 则

$$\mathbf{C} = (c_{ij})_{m \times s}, \quad c_{ij} = \sum_{k=1}^n a_{ik} \cdot b_{kj}$$

可以简单理解为 \mathbf{C} 的第 i 行第 j 列由 \mathbf{A} 的第 i 个行向量和 \mathbf{B} 的第 j 个列向量点积得到。

5. 方阵: n 行 n 列的矩阵称为 n 阶方阵。如常见的 n 阶单位矩阵:

$$E_n = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix}$$

6. 转置: 令 $A = (a_{ij})_{m \times n}$, 称 $B = (b_{ij})_{n \times m}$ 为 A 的转置, 记作 $B = A^T$, 其中 $b_{ij} = a_{ji}$ 。若 $A^T = A$, 则称 A 为对称方阵。

2.3.2 行列式

1. 二阶行列式:

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc$$

代表向量 (a, b) 和向量 (c, d) 所夹平行四边形的有向面积。

2. n 阶行列式: 称 n 阶方阵 A 的行列式为

$$\det A = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

3. 行列式的性质:

(1) 上三角矩阵的行列式为对角线乘积, 即

$$\det \mathbf{A} = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ 0 & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{vmatrix} = \prod_{i=1}^n a_{ii}$$

(2) 将方阵作转置, 行列式不变, 即 $\det \mathbf{A} = \det \mathbf{A}^T$;

(3) 若 \mathbf{A} 的某一行向量为 $\mathbf{0}$, 则 $\det \mathbf{A} = 0$;

(4) 互换 \mathbf{A} 的两行, $\det \mathbf{A}$ 变号;

(5) 将 \mathbf{A} 的某一行的 k 倍加到另一行, $\det \mathbf{A}$ 不变。

4. 行列式的求法: 按如下步骤进行。

(1) 若 $a_{11} = 0$, 即 \mathbf{A} 的第一列为 $\mathbf{0}$, 由 \mathbf{A}^T 有一行为 $\mathbf{0}$ 知, $\det \mathbf{A} = 0$;

(2) 否则, 若 $a_{11} = 0$, 则存在 $a_{j1} \neq 0$, 交换第 1 行和第 j 行后, $a_{11} \neq 0$, $\det \mathbf{A}$ 变号;

(3) 将第 i 行 ($i > 1$) 减去第一行的 $\frac{a_{i1}}{a_{11}}$ 倍, 之后 $a_{i1} = 0$;

(4) 对右下角的 $(n-1) \times (n-1)$ 方阵作相同处理, 返回第 (1) 步。

经上述处理后, 方阵 \mathbf{A} 被化为上三角方阵, 直接求对角元素乘积即可。

5. 行列式按行展开: 将 $\det(\mathbf{A})$ 按第 i 行展开, 有

$$\det \mathbf{A} = \sum_{j=1}^n a_{ij} A_{ij}$$

其中 $A_{ij} = (-1)^{i+j} M_{ij}$ 称为代数余子式, M_{ij} 为划去第 i 行和第 j 列后剩余部分的

行列式。同样可将 $\det \mathbf{A}$ 按第 j 列展开如下:

$$\det \mathbf{A} = \sum_{i=1}^n a_{ij} A_{ij}$$

2.3.3 在解线性方程组中的应用

1. 方程组的矩阵表示: 令系数矩阵、解向量、常数向量分别为

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}, \quad \mathbf{b} = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix}$$

则线性方程组可表示为

$$\mathbf{Ax} = \mathbf{b}$$

2. 克莱姆法则: 当 \mathbf{A} 为方阵时, 令 $d_i = \det(\alpha_1, \alpha_2, \cdots, \alpha_{i-1}, \mathbf{b}, \alpha_{i+1}, \cdots, \alpha_n)$, $d = \det \mathbf{A}$ 。若 $d \neq 0$, 则线性方程组有唯一解 $x_i = \frac{d_i}{d}$ 。

3. 线性方程组: 线性方程组有解判别定理: 方程组 $\mathbf{Ax} = \mathbf{b}$ 有解的充要条件是

$$\text{rank}(\alpha_1, \cdots, \alpha_n) = \text{rank}(\alpha_1, \cdots, \alpha_n, \mathbf{b})$$

即系数矩阵的秩等于增广矩阵的秩。

2.3.4 矩阵的性质

1. 矩阵的秩: 即列向量组 $(\alpha_1, \cdots, \alpha_n)$ 的秩。同一矩阵行向量组和列向量组的秩相等。

2, 矩阵乘积的性质: 当 A, B 均为 $n \times n$ 矩阵时, $|AB| = |A| \cdot |B|$; $\text{rank}(AB) \leq \text{rank} B$ 。

3. 退化方阵: 当 $\det(A) = 0$ 时, 称方阵 A 为退化的。

4. 矩阵的逆: 若存在方阵 B , 使得 $AB = BA = E$, 则称 B 为方阵 A 的逆矩阵, 记作 A^{-1} 。

5. 求矩阵的逆: 当 A 为非退化方阵时, 有两种方法求矩阵 A 的逆:

(1) 对矩阵 $[A|E]$ 作行变换, 得到 $[E|A^{-1}]$;

(2) 考虑伴随矩阵

$$A^* = \begin{pmatrix} A_{11} & A_{21} & \cdots & A_{n1} \\ A_{12} & A_{22} & \cdots & A_{n2} \\ \vdots & \vdots & & \vdots \\ A_{1n} & A_{2n} & \cdots & A_{nn} \end{pmatrix}$$

其中 A_{ij} 为 a_{ij} 位置的代数余子式。则有如下恒等式成立:

$$A^{-1} = \frac{1}{|A|} A^*$$

6. 初等矩阵: 由单位矩阵 E 经过一次初等变换得到的矩阵。

7. 矩阵等价: 若 A, B 为 $m \times n$ 矩阵, 且存在 m 阶可逆方阵 P 和 n 阶可逆方阵 Q , 使得 $A = PBQ$, 则称 A 和 B 等价, 此时 A 可由 B 经过初等变换得到。

2.4 二次型

2.4.1 二次型的矩阵表示

1. 二次型：一个关于 x_1, \dots, x_n 的二次齐次多项式

$$f(x_1, \dots, x_n) = a_{11}x_1^2 + 2a_{12}x_1x_2 + \dots + 2a_{n,n-1}x_nx_{n-1} + a_{nn}x_n^2$$

2. 矩阵表示：令对称矩阵

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix}, \quad \mathbf{x} = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

则上述二次型可用 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 表示。

3. 非退化线性替换：令 \mathbf{P} 为 n 阶可逆方阵，作非退化线性替换 $\mathbf{x} = \mathbf{P} \mathbf{y}$ ，则

$$\mathbf{x}^T \mathbf{A} \mathbf{x} = \mathbf{y}^T (\mathbf{P}^T \mathbf{A} \mathbf{P}) \mathbf{y}$$

4. 矩阵合同：若存在 n 阶可逆方阵 \mathbf{P} ，使得 $\mathbf{B} = \mathbf{P}^T \mathbf{A} \mathbf{P}$ ，则称矩阵 \mathbf{A} 和 \mathbf{B} 合同，此时二次型 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 可由非退化线性替换 $\mathbf{x} = \mathbf{P} \mathbf{y}$ 化为二次型 $\mathbf{y}^T \mathbf{B} \mathbf{y}$ 。

2.4.2 二次型的化简

1. 标准形：形如 $d_1x_1^2 + d_2x_2^2 + \dots + d_nx_n^2$ 的平方和形式。
2. 化简可行性：任一对称矩阵均合同于某一对角矩阵，即任意二次型均可通过非退化线性替换转化为标准形。
3. 配方法：

(1) 若存在 t 使得 $a_{tt} = 0$, 则按 x_t 为主元, 配方后换元以消去 x_t 的一次项;

(2) 若 $\forall t, a_{tt} = 0$, 但 $a_{ij} \neq 0$ 。令 $x_i = z_i + z_j, x_j = z_i - z_j$, 其余变量不变, 则 $a_{ij}x_ix_j$ 可转化为含 z_i^2 和 z_j^2 的部分, 返回 (1)。

4. 规范形: 形如 $z_1^2 + \cdots + z_p^2 - z_{p+1}^2 - \cdots - z_r^2$ 的二次型。任意实二次型可被非退化线性替换转化为唯一规范形。

5. 惯性指数: 正平方项的个数 p 称为正惯性指数, 负平方项的个数 $r - p$ 称为负惯性指数。

6. 正定二次型: $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x}^T \mathbf{A} \mathbf{x} > 0$, 则称二次型 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 正定。 n 元实二次型正定的充要条件是正惯性指数等于 n 。

7. 半正定: $\forall \mathbf{x} \in \mathbb{R}^n, \mathbf{x}^T \mathbf{A} \mathbf{x} \geq 0$, 则称二次型 $\mathbf{x}^T \mathbf{A} \mathbf{x}$ 半正定。

2.5 线性空间

2.5.1 线性空间的定义

1. 线性空间: 设 V 为非空集合, P 为数域, 在 V 上定义加法和数乘运算:

(1) 加法: $\gamma = \alpha + \beta$, 其中 $\alpha, \beta \in V$ 。

(2) 数乘: $\delta = k\alpha$, 其中 $k \in P, \alpha \in V$ 。

且加法具有结合律、交换律、单位元 (0) 和逆元 (减法); 数乘具有结合律, 单位元 (1); 加法和数乘满足两项分配律, 则称 V 为数域 P 上的线性空间, 将 V 中的元素称为向量。

2. 线性空间举例: 分量属于 P 的全体 n 元数组 P^n ; 数域 P 上的一元多项式环 $P[x]$ 。

2.5.2 基与坐标

1. 可继承 R^n 中向量的概念：线性表示、线性相关、线性无关、线性组合、极大线性无关组、秩。由于之前已经讨论过这些概念，此处不再赘述。

2. 维数：若 V 中最多存在 n 个线性无关的向量，则称 V 为 n 维线性空间。

3. 基： n 维线性空间 V 中， n 个线性无关的向量 $\varepsilon_1, \dots, \varepsilon_n$ 称为 V 的一组基。使用 $\varepsilon_1, \dots, \varepsilon_n$ 可以线性表示 V 中的所有向量。

4. 坐标：令 $\alpha \in V$ ，则 α 可唯一地表示为如下形式：

$$\alpha = x_1 \varepsilon_1 + \dots + x_n \varepsilon_n$$

称 (x_1, \dots, x_n) 为 α 在基 $\varepsilon_1, \dots, \varepsilon_n$ 下的坐标。

5. 向量的基-坐标表示法：

$$\alpha = (\varepsilon_1, \dots, \varepsilon_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

2.5.3 基变换与坐标变换

1. 两组基的转换：令 $\varepsilon_1, \dots, \varepsilon_n$ 和 $\varepsilon'_1, \dots, \varepsilon'_n$ 为 V 的两组基，且

$$\varepsilon'_i = a_{1i} \varepsilon_1 + \dots + a_{ni} \varepsilon_n = (\varepsilon_1, \dots, \varepsilon_n) \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{pmatrix}$$

则有如下等式成立，其中 $A = (a_{ij})_{n \times n}$ 称为 $\varepsilon_1, \dots, \varepsilon_n$ 到 $\varepsilon'_1, \dots, \varepsilon'_n$ 的过渡矩阵。

$$(\varepsilon'_1, \dots, \varepsilon'_n) = (\varepsilon_1, \dots, \varepsilon_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

2. 同一向量在两组基下的坐标变换：令

$$x = (\varepsilon_1, \dots, \varepsilon_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

代入 $(\varepsilon'_1, \dots, \varepsilon'_n) = (\varepsilon_1, \dots, \varepsilon_n)A$ ，得到

$$x = (\varepsilon'_1, \dots, \varepsilon'_n)A^{-1} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}$$

因此 x 在基 $\varepsilon'_1, \dots, \varepsilon'_n$ 下的坐标为 $A^{-1}(x_1, \dots, x_n)^T$ 。

2.5.4 线性子空间

1. 定义：若 $W \subset V$, $W \neq \Phi$ ，且 W 对加法和数乘运算封闭，则称 W 为 V 的线性子空间。

2. 生成子空间：设 $\alpha, \dots, \alpha_m \in V$ 为一组线性无关的向量，则所有能被 $\alpha_1, \dots, \alpha_m$ 线性表示的向量构成 V 的一个子空间，称为由 $\alpha_1, \dots, \alpha_m$ 生成的子空间，记作 $L(\alpha_1, \dots, \alpha_m)$ 。

3. 子空间基的扩充定理*：若 $\alpha_1, \dots, \alpha_m$ 为子空间 W 的一组基，则必存在 $\alpha_1, \dots, \alpha_m, \alpha_{m+1}, \dots, \alpha_n$ 是 V 的一组基。

4. 子空间的交：若 $V_1 \subset V, V_2 \subset V$ 均为 V 的子空间，则 $V_1 \cap V_2$ 也为 V 的子空间。

5. 子空间的和：若 $V_1 \subset V, V_2 \subset V$ 均为 V 的子空间，则 $\{\alpha_1 + \alpha_2 | \alpha_1 \in V_1, \alpha_2 \in V_2\}$ 也为 V 的子空间，记作 $V_1 + V_2$ 。

6. 维数公式*：记 $\dim V$ 为 V 的维数，则有如下公式成立：

$$\dim V_1 + \dim V_2 = \dim(V_1 + V_2) + \dim(V_1 \cap V_2)$$

7. 子空间的直和：若 $V_1 \cap V_2 = \{0\}$ ，则 V_1 和 V_2 的和被称为直和，记作 $V_1 \oplus V_2$ 。

8. 直和的性质*： $V_1 + V_2 = V_1 \oplus V_2 \Leftrightarrow \dim(V_1 + V_2) = \dim V_1 + \dim V_2$ 。

2.6 线性变换

2.6.1 映射与变换

1. 映射：某一对应规则 $f: A \rightarrow B$ ，使得 $\forall x \in A, \exists! y \in B, y = f(x)$ 。此时称 y 为 x 在映射 f 下的像， x 称为原像。

2. 变换：称集合 A 到自身的映射 $f: A \rightarrow A$ 为 A 到自身的变换。

3. 线性变换：保持加法和数乘运算的变换，用 \mathcal{A}, \mathcal{B} 等花体字母表示，即

$$\mathcal{A}(\alpha + \beta) = \mathcal{A}\alpha + \mathcal{A}\beta; \mathcal{A}(k\alpha) = k \cdot \mathcal{A}\alpha$$

2.6.2 线性变换的运算

1. 线性变换的乘法: $\mathcal{A}\mathcal{B}(\alpha) = \mathcal{A}[\mathcal{B}\alpha]$ 。乘法满足结合律, 但一般不满足交换律。
2. 线性变换的加法: $(\mathcal{A} + \mathcal{B})\alpha = \mathcal{A}\alpha + \mathcal{B}\alpha$ 。加法满足交换律和结合律。
3. 逆变换: 若 $\mathcal{A}\mathcal{B} = \mathcal{B}\mathcal{A} = \mathcal{E}$, 其中 \mathcal{E} 为恒等变换, 则称 \mathcal{B} 为 \mathcal{A} 的逆变换。

2.6.3 线性变换的矩阵表示

1. 线性变换与向量的对应关系*: 设 $\varepsilon_1, \dots, \varepsilon_n$ 为 V 的一组基, 则 $\forall \alpha_1, \dots, \alpha_n \in V$, $\exists ! \mathcal{A}$, 使得 $\mathcal{A}\varepsilon_i = \alpha_i$ 。
2. 矩阵表示: 将 $\mathcal{A}\varepsilon_i$ 用 $\varepsilon_1, \dots, \varepsilon_n$ 线性表示:

$$\mathcal{A}\varepsilon_i = a_{1i}\varepsilon_1 + a_{2i}\varepsilon_2 + \cdots + a_{ni}\varepsilon_n = (\varepsilon_1, \dots, \varepsilon_n) \begin{pmatrix} a_{1i} \\ a_{2i} \\ \vdots \\ a_{ni} \end{pmatrix}$$

则有

$$\mathcal{A}(\varepsilon_1, \dots, \varepsilon_n) = (\mathcal{A}\varepsilon_1, \dots, \mathcal{A}\varepsilon_n) = (\varepsilon_1, \dots, \varepsilon_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix}$$

将矩阵 $A = (a_{ij})_{n \times n}$ 称为 \mathcal{A} 在基 $\varepsilon_1, \dots, \varepsilon_n$ 下的矩阵, 记作

$$\mathcal{A}(\varepsilon_1, \dots, \varepsilon_n) = (\varepsilon_1, \dots, \varepsilon_n)A$$

3. 同一线性变换在不同基下的矩阵: 设 $\mathcal{A}(\varepsilon_1, \dots, \varepsilon_n) = (\varepsilon_1, \dots, \varepsilon_n)A$, 同时有 $\mathcal{A}(\varepsilon'_1, \dots, \varepsilon'_n) = (\varepsilon'_1, \dots, \varepsilon'_n)B$, 而 $(\varepsilon_1, \dots, \varepsilon_n) = (\varepsilon'_1, \dots, \varepsilon'_n)P$, 则第一个等式两边可分别化为

$$\begin{aligned}\mathcal{A}(\varepsilon_1, \dots, \varepsilon_n) &= \mathcal{A}(\varepsilon'_1, \dots, \varepsilon'_n)P = (\varepsilon'_1, \dots, \varepsilon'_n)BP \\ (\varepsilon_1, \dots, \varepsilon_n)A &= (\varepsilon'_1, \dots, \varepsilon'_n)PA\end{aligned}$$

得到 $BP = PA$, 即 $A = P^{-1}BP$ 。

4. 相似矩阵: 设 A, B 均为 n 阶方阵。若存在可逆的 n 阶方阵 P , 使 $A = P^{-1}BP$, 则称矩阵 A 与矩阵 B 相似, 记作 $A \sim B$ 。相似矩阵的本质是同一线性变换在不同基下的矩阵。

2.6.4 特征值与特征向量

1. 定义: 令 \mathcal{A} 是 V 上的线性变换, 若存在 $\lambda \in P$ 和 $\xi \in V$, 使得

$$\mathcal{A}\xi = \lambda\xi$$

则称 λ 为 \mathcal{A} 的一个特征值, ξ 为 \mathcal{A} 属于特征值 λ 的特征向量。

2. 求解: 设 \mathcal{A} 在某组基下的矩阵为 A , 解 A 的特征多项式的所有根, 即关于 λ 的方程:

$$|\lambda E - A| = 0$$

解出所有符合条件的 λ 并逐一代入，解线性方程组：

$$(\lambda E - A)x = 0$$

解得的基础解系便是属于 λ 的特征向量。

3. 性质*：相似的矩阵有相同的特征值，即 $A = P^{-1}BP$ 时， $|\lambda E - A| = |\lambda E - B|$ 。换句话说，同一线性变换在不同基下的特征值不变。

4. Hamilton-Cayley 定理，记 A 的特征多项式为 $f(\lambda) = |\lambda E - A|$ ，则 $f(A) = 0$ 。

5. 对角矩阵的转化：若 \mathcal{A} 存在 n 个线性无关的特征向量，则 \mathcal{A} 在某组基下的矩阵是对角形的。显然若 \mathcal{A} 的特征多项式有 n 个不同的根，则 \mathcal{A} 在当前基下的矩阵可相似对角化。

6. 相似对角化方法：令对角矩阵

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix}$$

对应特征值 λ_i 的特征向量为 ξ_i ，则令矩阵 $P = (\xi_1, \cdots, \xi_n)$ ，则有 $AP = P\Lambda$ ，即 $P^{-1}AP = \Lambda$ 。

2.6.5 线性变换的值域与核

1. 定义：令 $\{y | \exists x \in V, \mathcal{A}x = y\}$ 为线性变换 \mathcal{A} 的值域，记作 $\mathcal{A}V$ 或 $\text{Im}\mathcal{A}$ ；令 $\{x \in V | \mathcal{A}x = 0\}$ 为线性变换 \mathcal{A} 的核，记作 $\mathcal{A}^{-1}(0)$ 或 $\ker\mathcal{A}$ 。

2. 秩和零度： $\mathcal{A}V$ 的维数称为 \mathcal{A} 的秩； $\mathcal{A}^{-1}(0)$ 的维数称为 \mathcal{A} 的零度。

3. 重要定理*：令 $\mathcal{A}(x_1), \cdots, \mathcal{A}x_m$ 为 $\mathcal{A}V$ 的一组基， x_{m+1}, \cdots, x_n 为 $\mathcal{A}^{-1}(0)$

的一组基, 则 $\mathbf{x}_1, \cdots, \mathbf{x}_m, \mathbf{x}_{m+1}, \cdots, \mathbf{x}_n$ 为 V 的一组基, 由此可得

$$\dim(\mathcal{A}V) + \dim(\mathcal{A}^{-1}(\mathbf{0})) = \dim V$$

2.6.6 不变子空间

1. 定义: 设 $W \subset V$, 若 $\forall \xi \in W$, 均有 $\mathcal{A}\xi \in W$, 则称 W 为 \mathcal{A} 的不变子空间。

2. 线性空间的分解: 设 \mathcal{A} 的特征多项式 $f(\lambda)$ 可分解为一次因式的乘积:

$$f(\lambda) = \prod_{i=1}^s (\lambda - \lambda_i)^{r_i}$$

则 V 可分解为若干不变子空间的直和 $V = V_1 \oplus V_2 \cdots \oplus V_s$, 其中 $V_i = \{\xi \in V | (\mathcal{A} - \lambda_i \mathcal{E})^{r_i} \xi = 0\}$, 该定理用于将矩阵分解为分块准对角形。

2.7 λ -矩阵

2.7.1 基本概念

1. 定义: 若矩阵 A 中的元素是 λ 的多项式, 则称 A 为 λ -矩阵, 记作 $A(\lambda)$ 。
2. 可逆: 若 $\exists B(\lambda)$, 使得 $A(\lambda)B(\lambda) = B(\lambda)A(\lambda) = E$, 则称 $B(\lambda)$ 是 $A(\lambda)$ 的逆矩阵, 记作 $A^{-1}(\lambda)$ 。
3. 可逆的充要条件: $A(\lambda)$ 可逆 $\Leftrightarrow |A(\lambda)|$ 为非零常数。

2.7.2 标准形

1. 初等变换: 有以下三种形式:

- (1) 互换两行/两列的位置;
 - (2) 某行/列乘以非零常数 c ;
 - (3) 某行/列加上另一行/列的 $\varphi(\lambda)$ 倍。
2. 等价: 若 $A(\lambda)$ 可由一系列初等变换得到 $B(\lambda)$, 则称 $A(\lambda)$ 与 $B(\lambda)$ 等价。
3. 标准形: 任意 $s \times n$ 的 λ -矩阵 $A(\lambda)$ 均可化为下述形式的矩阵:

$$\begin{pmatrix} d_1(\lambda) & 0 & 0 & \cdots & 0 & 0 \\ 0 & d_2(\lambda) & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & d_r(\lambda) & \cdots & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 0 \end{pmatrix}$$

其中 $r \geq 1$, $d_i(\lambda)$ 为首一多项式, 且 $d_i(\lambda) \mid d_{i+1}(\lambda)$ 。该形式称为 λ -矩阵的标准形。

4. 化为标准形的步骤:

- (1) 调换行、列, 使 $a_{11}(\lambda)$ 是其它所有元素的公因子;
- (2) 从 a_{11} 出发, 通过初等变换, 使得 a_{1j} 和 a_{i1} 均化为 0;
- (3) 对右下角 $(n-1) \times (n-1)$ 的矩阵块重复上述操作。

2.7.3 相似不变量

1. 不变因子: 标准形中主对角线上非零元素 $d_1(\lambda), d_2(\lambda), \dots, d_r(\lambda)$ 称为 $A(\lambda)$ 的不变因子。
2. 初等因子: 将所有不变因子分解为一次首一多项式的乘积, 所有一次因式的方幂称为矩阵 $A(\lambda)$ 的初等因子。

3. 行列式因子： $A(\lambda)$ 中全部 k 阶子式的首一最大公因式 $D_k(\lambda)$ 称为 $A(\lambda)$ 的 k 阶行列式因子。有如下等式成立：

$$D_k(\lambda) = d_1(\lambda)d_2(\lambda) \cdots d_k(\lambda)$$

2.7.4 数字矩阵的相似标准形

1. 数字矩阵相似的条件： $A \sim B \Leftrightarrow \lambda E - A$ 和 $\lambda E - B$ 等价。
2. 若尔当标准形：形为

$$J_i(\lambda) = \begin{pmatrix} \lambda & 0 & \cdots & 0 & 0 \\ 1 & \lambda & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda & 0 \\ 0 & 0 & \cdots & 1 & \lambda \end{pmatrix}$$

的矩阵称为若尔当块。将由若干若尔当块组成的矩阵

$$\begin{pmatrix} J_1 & 0 & \cdots & 0 \\ 0 & J_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \vdots & J_n \end{pmatrix}$$

称为若尔当形矩阵。

3. 矩阵化为若尔当标准形：设 $\lambda E - A$ 的初等因子为 $(\lambda - \lambda_1)^{k_1}, \dots, (\lambda - \lambda_s)^{k_s}$ ，则每个初等因子 $(\lambda - \lambda_i)^{k_i}$ 对应一个若尔当块 $J_i(\lambda_i)$ ，将所有 J_i 组合便得到 A 的

若尔当标准形, 且除了若尔当块的排列方式以外, 该标准形唯一。

4. 友矩阵: 设多项式 $d(\lambda) = \lambda^n + a_1\lambda^{n-1} + \cdots + a_n$, 则称

$$A = \begin{pmatrix} 0 & 0 & \cdots & 0 & -a_n \\ 1 & 0 & \cdots & 0 & -a_{n-1} \\ 0 & 1 & \cdots & 0 & -a_{n-2} \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & -a_1 \end{pmatrix}$$

为 $d(\lambda)$ 的友矩阵。

5. 有理标准形矩阵: 称准对角矩阵

$$A = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & A_n \end{pmatrix}$$

为有理标准形矩阵, 其中 A_i 为 $d_i(\lambda)$ 的友矩阵, 且 $d_i(\lambda) \mid d_{i+1}(\lambda)$ 。

6. 矩阵的有理标准形: 令 $\lambda E - A$ 的不变因子为 $1, 1, \cdots, d_1(\lambda), \cdots, d_s(\lambda)$, 且 A_i 为 $d_i(\lambda)$ 的友矩阵, 便得到矩阵 A 的有理标准形。

2.8 欧几里得空间

2.8.1 欧氏空间与内积

1. 内积: 设 V 是 \mathbb{R} 上的线性空间, 在 V 上定义二元实函数 (α, β) , 满足如下条件:

(1) 正定性: $(\alpha, \alpha) \geq 0$, 当且仅当 $\alpha = \mathbf{0}$ 时取等;

(2) 对称性: $(\alpha, \beta) = (\beta, \alpha)$;

(3) 线性性: $(k_1\alpha + k_2\beta, \gamma) = k_1(\alpha, \gamma) + k_2(\beta, \gamma)$ 。则称 (α, β) 为 V 上的内积, 定义了内积的线性空间称为欧几里得空间。

2. 欧氏空间 \mathbb{R}^n : 在 \mathbb{R}^n 上, 令 $\alpha = (a_1, \dots, a_n)$, $\beta = (b_1, \dots, b_n)$, 定义如下内积:

$$(\alpha, \beta) = a_1b_1 + \dots + a_nb_n$$

则 \mathbb{R}^n 构成一最常用的欧氏空间, 下文若无特殊说明均在此空间上讨论。

3. Cauchy-Schwarz 不等式: $|(\alpha, \beta)| \leq |\alpha| \cdot |\beta|$, 当 $\alpha = k\beta$ 时取等。此不等式对任一符合定义的内积均适用。

4. 向量夹角: 规定向量 α 和 β 的夹角为

$$\langle \alpha, \beta \rangle = \arccos \frac{(\alpha, \beta)}{|\alpha| \cdot |\beta|}$$

5. 模: 定义 $\|\alpha\| = (\alpha, \alpha)$ 为向量 α 的模, 又称为 α 的长度。

6. 正交: 若 $(\alpha, \beta) = 0$, 则称 α 和 β 正交, 记作 $\alpha \perp \beta$ 。

7. 度量矩阵: 令 $\varepsilon_1, \dots, \varepsilon_n$ 为 \mathbb{R}^n 上的一组基, $\mathbf{x} = x_1\varepsilon_1 + \dots + x_n\varepsilon_n$, $\mathbf{y} = y_1\varepsilon_1 + \dots + y_n\varepsilon_n$, 则有

$$(\mathbf{x}, \mathbf{y}) = \sum_{i=1}^n \sum_{j=1}^n x_i y_j (\varepsilon_i, \varepsilon_j) = \mathbf{x}^T \mathbf{A} \mathbf{y}$$

其中方阵

$$A = \begin{pmatrix} (\varepsilon_1, \varepsilon_1) & (\varepsilon_1, \varepsilon_2) & \cdots & (\varepsilon_1, \varepsilon_n) \\ (\varepsilon_2, \varepsilon_1) & (\varepsilon_2, \varepsilon_2) & \cdots & (\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & & \vdots \\ (\varepsilon_n, \varepsilon_1) & (\varepsilon_n, \varepsilon_2) & \cdots & (\varepsilon_n, \varepsilon_n) \end{pmatrix}$$

称为度量矩阵。

2.8.2 标准正交基

1. 正交向量组：若 $\alpha_1, \dots, \alpha_m$ 中， $\alpha_i \perp \alpha_j$ 对 $\forall i \neq j$ 成立，则称 $\alpha_1, \dots, \alpha_m$ 为正交向量组。注意正交向量组总是线性无关的。

2. 标准正交基：若一组基 $\varepsilon_1, \dots, \varepsilon_n$ 两两正交，且 $|\varepsilon_i| = 1$ 恒成立，则称 $\varepsilon_1, \dots, \varepsilon_n$ 为标准正交基。

3. 标准正交基的性质：度量矩阵为单位矩阵，即 $(x, y) = x^T y$ 。

4. 施密特正交化：已知 $\alpha_1, \dots, \alpha_n$ 为 \mathbb{R}^n 上一组基，取

$$\begin{aligned} \beta_1 &= \alpha_1, \beta_2 = \alpha_2 - \frac{(\alpha_2, \beta_1)}{(\beta_1, \beta_1)} \beta_1, \dots, \\ \beta_n &= \alpha_n - \frac{(\alpha_n, \beta_1)}{(\beta_1, \beta_1)} \beta_1 - \dots - \frac{(\alpha_n, \beta_{n-1})}{(\beta_{n-1}, \beta_{n-1})} \beta_{n-1} \end{aligned}$$

则 $\frac{\beta_1}{|\beta_1|}, \dots, \frac{\beta_n}{|\beta_n|}$ 为一组标准正交基。

2.8.3 正交变换

1. 定义：保持向量内积不变的变换：

$$(\mathcal{A}\alpha, \mathcal{A}\beta) = (\alpha, \beta)$$

2. 性质*：正交变换保持向量长度不变；在任一组标准正交基下的矩阵是正交矩阵。

3. 正交矩阵：满足 $\mathbf{A}\mathbf{A}^T = \mathbf{E}$ 的矩阵。注意标准正交基构成的矩阵 $(\boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_n)$ 为正交矩阵。

4. 正交子空间：若 $\forall \boldsymbol{\alpha} \in V_1, \boldsymbol{\beta} \in V_2$ ，均有 $(\boldsymbol{\alpha}, \boldsymbol{\beta}) = 0$ ，则称 V_1 和 V_2 正交。

5. 正交补：若 $V_1 \perp V_2$ ，且 $V_1 + V_2 = V$ ，则称 V_1 为 V_2 的正交补。

2.8.4 实对称矩阵的正交对角化

1. 目标：对实对称矩阵 \mathbf{A} ，寻找正交矩阵 \mathbf{T} ，使得 $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ 为对角矩阵。

2. 步骤：

(1) 求出实对称矩阵 \mathbf{A} 的特征值；

(2) 对每个特征值，求解对应特征向量；

(3) 将特征向量施密特正交化并单位化，排列为正交矩阵 \mathbf{T} ，此时 $\mathbf{T}^{-1}\mathbf{A}\mathbf{T}$ 正好为对角矩阵，且对角元素是矩阵 \mathbf{A} 的所有特征值，按特征向量的顺序排列。

第三章 概率论与数理统计

Mathematical Analysis

概率论与数理统计是由数分高代派生出来的应用学科,用于刻画日常生活中随机发生的事件,具有很高的应用价值. 其中,概率论主要研究随机变量的分布与特征,而数理统计主要研究通过样本对未知分布进行估计.

概率论的重点: 概率的定义, 条件概率与独立性, 一元或多元随机变量分布, 常用分布函数, 随机变量的特征数, 大数定律和中心极限定理

数理统计的重点: 基本概念与三大分布, 参数估计, 假设检验, 方差分析, 回归分析

3.1 随机事件与概率

之前数学分析研究的内容都是具有确定解析式或约束条件的函数, 但概率论引进了随机因素, 即实验和结果并不是一一对应的, 一次实验可能会出多种结果. 这一部分的任务是使用概率这一量化方式, 将随机性规范化.

3.1.1 随机事件

1. **随机现象:** 重复实验会出现不同结果的现象.
2. **样本空间:** 随机现象可能出现的结果组成的集合.
3. **随机事件:** 样本空间的子集. 当实验结果属于此子集时, 称随机事件发生.
4. **随机变量:** 用于描述随机事件的人为设定变量 (非正式定义).
5. **事件的运算:** 和集合一致, 有交并补余四大运算. 有两个公式很重要.
 - (1) 集合减法公式: $A - B = A \cap \bar{B}$.
 - (2) 德摩根律: $\overline{A \cup B} = \bar{A} \cap \bar{B}; \overline{A \cap B} = \bar{A} \cup \bar{B}$.
6. **事件域:** 令 Ω 为样本空间, 定义事件域 \mathcal{F} 符合下列性质:
 - (1) $\Omega \in \mathcal{F}$; (2) $A \in \mathcal{F} \Rightarrow \bar{A} \in \mathcal{F}$; (3) $A_n \in \mathcal{F} \Rightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{F}$.

3.1.2 概率

1. **公理化定义:** 在事件域 (Ω, \mathcal{F}) 上定义可测函数 $P(A)$ 满足:
 - (1) 非负性: $P(A) \geq 0$; (2) 正则性: $P(\Omega) = 1$;
 - (3) 可列可加性: 事件 A_1, \dots, A_n 互不相容时, $P(\bigcup_{n=1}^{\infty} A_n) = \sum_{n=1}^{\infty} P(A_i)$.
2. **用频率定义概率:** 令 $n(A)$ 为事件 A 发生的频数, 则可用大量重复事件的频率表示概率: $P(A) = \lim_{n \rightarrow \infty} \frac{n(A)}{n}$.

3. 古典概型: 若样本空间有 n 个等可能发生的样本点, 则事件 A 包含 k 个样本点时, $P(A) = \frac{k}{n}$.

4. 几何概型: 若样本空间 Ω 的面积测度为 S_n , 事件 A 包含其中面积为 S_A 的一部分, 则 $P(A) = \frac{S_A}{S_n}$. (蒙特卡罗法的理论依据)

5. 贝叶斯概率: 对事件发生可能性的主观预测, 在机器学习中使用频率很高.

3.1.3 概率的性质

1. 有限可加性: 若 A_1, \dots, A_n 互不相容, 则 $P(\bigcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$.
2. 单调性: 若 $A \subset B$, 则 $P(A) \leq P(B)$.
3. 加法公式: $P(A \cup B) = P(A) + P(B) - P(AB)$.

3.1.4 条件概率

1. 定义: $P(A|B)$ 表示已知 B 发生的条件下 A 发生的概率. $P(A|B) = \frac{P(AB)}{P(B)}$.
2. 乘法公式: $P(AB) = P(B)P(A|B)$, 即定义式的变种.
3. 全概率公式: 若 B_i 互不相容, 且 $\bigcup_{i=1}^n B_i = \Omega$, 则

$$P(A) = \sum_{i=1}^n P(A|B_i)P(B_i)$$

4. 贝叶斯公式: 用先验概率推后验概率. 若 B_i 互不相容, 且 $\bigcup_{i=1}^n B_i = \Omega$, 则

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

5. 独立性: 若 $P(A|B) = P(A)$, 即 $P(AB) = P(A)P(B)$, 则称事件 A 和 B 相互独立.

3.2 随机变量及其分布

用概率描述随机事件发生可能性的大小后, 为了更充分认识随机事件, 我们引入随机变量来刻画随机事件, 如抽奖是随机事件, 在此基础上可以定义随机变量“是否中奖”, 这是一个二值随机变量 (0/1).

使用随机变量来描述随机事件, 能更方便地研究随机事件中我们感兴趣的性质, 比如随机变量“灯泡坏掉的个数”能帮助我们衡量灯泡的寿命. 这些随机变量取值的规律可以用分布来描述, 离散随机变量和连续随机变量的刻画方式略有区别.

3.2.1 随机变量

- 1. 定义:** 样本空间 Ω 上的实值函数 $X(\omega)$.
- 2. 离散随机变量的确定:** 使用分布列描述.

X	X_1	X_2	\cdots	X_n
P	p_1	p_2	\cdots	p_n

其中 p_i 表示随机变量 X 取值 X_i 的概率, $\sum_{i=1}^n p_i = 1$.

- 3. 连续随机变量的描述:** 使用分布函数与概率密度函数.

(1) 分布函数 $F(x)$: $F(x) = P(X \leq x)$, 是单调递增的右连续函数, 且 $F(+\infty) = 1, F(-\infty) = 0$.

(2) p.d.f 概率密度函数 $p(x)$: $p(x) = F'(x)$, 是非负函数且 $\int_{-\infty}^{+\infty} p(x)dx = 1$.

3.2.2 常用分布及概率密度函数

1. 离散分布

(1) 泊松分布: $P(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}$, 用于计数过程, 记作 $X \sim P(\lambda)$.

(2) 伯努利分布: $P(X = 1) = p, P(X = 0) = 1 - p$, 又称两点分布.

(3) 二项分布: $P(X = k) = \binom{n}{k} p^k (1 - p)^{n-k}$, 即 n 重伯努利分布中事件发生的次数, 记作 $X \sim b(n, p)$.

(4) 几何分布: $P(X = k) = (1 - p)^{k-1} p$, 具有无记忆性.

2. 连续分布

(1) 正态分布: $p(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$, 是最常用的分布. 记作 $X \sim N(\mu, \sigma^2)$, 标准正态分布即 $N(0, 1)$.

(2) 均匀分布: $p(x) = \frac{1}{b-a}$, 其中 $x \in (a, b)$, 记作 $X \sim U(a, b)$.

(3) 指数分布: $p(x) = \lambda e^{-\lambda x}$, 其中 $x \geq 0$, 记作 $X \sim \epsilon(\lambda)$, 具有无记忆性.

(4) 伽马分布: $p(x) = \frac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\lambda x}$, 其中 $x \geq 0$, 记作 $X \sim Ga(\alpha, \lambda)$. 特殊地, $Ga(\frac{n}{2}, \frac{1}{2}) = \chi^2(n)$ 为卡方分布, 统计中常用.

3.2.3 数字特征

1. 数学期望: X 在不同取值数按概率的加权平均数, 是消除随机性的主要手段, 记作 EX . 在离散场合, $EX = \sum_{i=1}^{\infty} p_i x_i$. 在连续场合, $EX = \int_{-\infty}^{+\infty} x p(x) dx$.

2. 方差: $DX = E[(X - EX)^2]$, 也记作 $\text{Var}(X)$, 用于衡量数据的集中程度. 常用的计算公式为

$$DX = E(X^2) - (EX)^2$$

3. 标准差: $\sigma(x) = \sqrt{DX}$, 也记作 $\text{Std}(X)$, 好处是与 X 的量纲一致.

4. 切比雪夫不等式: $P(|X - EX| \geq \varepsilon) \leq \frac{DX}{\varepsilon^2}$.

3.2.4 随机变量函数的分布

1. 离散情形: 先求各项的像 $g(x_1), \dots, g(x_n)$, $g(x_i)$ 对应概率仍为 p_i , 再合并相同项.

2. 连续情形: 若 $Y = g(x)$ 严格单调, 反函数为 $x = h(y)$, X 的概率密度函数为 $p(x)$, 则 Y 的概率密度函数为 $p_Y(y) = p_X(h(y)) \cdot |h'(y)|$. 一般情况下, 需要根据 $P(g(x) \leq y)$ 反解出 x 的范围, 再利用 X 的分布函数求解.

3.3 多元随机变量及其分布

若样本点含有不止一个我们感兴趣的属性, 如身体指标包含身高和体重, 则可定义多元随机变量来刻画这些指标的分布. 研究多元随机变量, 除了明确各分量的分布外, 还需要研究各分量间的相关关系, 以及给定某条件后的分布情况.

事实上, 只要给定多元随机变量的联合分布, 就能得到所有信息, 该部分的目的就是掌握将信息从联合分布中提取出来的方法.

3.3.1 多元随机变量

1. 定义: 样本空间 Ω 上的向量值函数 $X(\omega) = (X_1(\omega), \dots, X_n(\omega))$.
2. 联合分布函数: $F(x_1, \dots, x_n) = P(X_1 \leq x_1, \dots, X_n \leq x_n)$
3. 离散情形的联合分布列: 仅用于二元分布 (X, Y) , 用 i 行 j 列元素 p_{ij} 表示 $X = X_i, Y = Y_j$ 的概率, 其中 $\sum_{i,j} p_{ij} = 1$.
4. 连续情形的联合密度函数: $p(x_1, \dots, x_n) = \frac{\partial^n F(x_1, \dots, x_n)}{\partial x_1 \partial x_2 \dots \partial x_n}$.
5. 多元正态分布: 最重要的多元连续分布. 令 $x = (x_1, \dots, x_n)$, 均值向量为 μ ,

协方差矩阵为 Γ , 则 n 元正态分布的联合密度函数

$$p(x_1, \cdots, x_n) = \frac{1}{(2\pi)^{\frac{n}{2}} |\Gamma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T \Gamma^{-1}(x-\mu)}$$

特殊地, 当 $n = 2$ 时, 二元正态分布为

$$p(x, y) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}[(\frac{x-\mu_1}{\sigma_1})^2 - 2\rho\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + (\frac{y-\mu_2}{\sigma_2})^2]}$$

记作 $(X, Y) \sim N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

3.3.2 边缘分布

1. 边缘分布函数: $F_x(x) = F(x, \infty)$, $F_y(y) = F(\infty, y)$.
2. 离散情形的边缘分布列: $P(X = X_i) = \sum_j P(X = X_i, Y = Y_j)$; $P(Y = Y_j) = \sum_i P(X = X_i, Y = Y_j)$
3. 连续情形的边缘密度函数: $p_X(x) = \int_{-\infty}^{\infty} p(x, y)dy$; $p_Y(y) = \int_{-\infty}^{\infty} p(x, y)dx$.
4. 随机变量的独立性: 若 $\prod_{i=1}^n p_i(X_i) = p(x_1, \cdots, x_n)$, 即联合密度函数为边缘密度函数之积, 则称 X_1, \cdots, X_n 相互独立.

3.3.3 多元随机变量函数的分布

1. $Z = X + Y$ 的分布: 可用后面提到的特征函数法, 也可用卷积公式, 即 $p_Z(z) = \int_{-\infty}^{\infty} p_X(x)p_Y(z-x)dx$.
2. 次序统计量分布: 若 $X_{(1)}, \cdots, X_{(n)}$ 独立同分布且升序排列, 则第 k 个次序统计量 $X_{(k)}$ 的概率密度函数为

$$p_{(k)}(x) = \frac{n!}{(k-1)!(n-k)!} [F(x)]^{k-1} p(x) [1-F(x)]^{n-k}$$

特殊地, $\min X$ 即 $X_{(1)}$ 的概率密度函数为 $n[1 - F(x)]^{n-1}p(x)$;

$\max X$ 即 $X_{(n)}$ 的概率密度函数为 $n[F(x)]^{n-1}p(x)$.

3. 变量变换法: 令 $u = u(x, y), v = v(x, y)$, 从中反解出 $x = x(u, v), y = y(u, v)$, 则 $p(u, v) = p(x, y) \left| \frac{\partial(x, y)}{\partial(u, v)} \right|$.

3.3.4 多元随机变量的特征数

1. 数学期望: $g(x, y)$ 的期望为 $\int_R g(x, y)p(x, y)dxdy$.

2. 方差: 定义不变, 仍有 $\text{Var}(x) = E(X - EX), \text{Var}(y) = E(Y - EY)$.

3. 协方差: $\text{Cov}(X, Y) = E[(X - EX)(Y - EY)] = E(XY) - EXEY$, 用于刻画两变量的相关程度.

4. 相关系数: $\text{Corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$. 当 $\text{Corr}(x) \in (0, 1]$ 时, 称 X 和 Y 正相关; $\text{Corr}(x) \in [-1, 0)$ 时, 称 X 和 Y 负相关; $\text{Corr}(x) = 0$ 时, 称 X 和 Y 不相关.

5. 方差运算性质: $\text{Var}(X \pm Y) = \text{Var}(X) + \text{Var}(Y) \pm 2\text{Cov}(X, Y)$

6. n 元随机变量的协方差矩阵:

$$\Gamma = \begin{pmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \cdots & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Cov}(X_n, X_n) \end{pmatrix}$$

用于刻画各分量之间的总体相关性.

3.3.5 条件分布与条件期望

1. 离散条件分布: $P_{ij} = P(X = X_i | Y = y_j) = \frac{p_{ij}}{\sum_i p_{ij}}$.

2. 连续条件分布: $p(y|x) = \frac{p(x, y)}{p_X(x)}$; $p(x|y) = \frac{p(x, y)}{p_Y(y)}$

3. 全概率公式: $p_Y(y) = \int_{-\infty}^{\infty} p_X(x)p(y|x)dx$.

4. 贝叶斯公式:

$$p(x|y) = \frac{p(y|x)p_X(x)}{\int_{-\infty}^{\infty} p(y|x)p_X(x)dx}$$

5. 条件数学期望: $E(X|Y = y) = \int_{-\infty}^{\infty} xp(x|y)dx$.

6. 重期望公式: $E[E(X|Y)] = EX$.

3.4 大数定律和中心极限定理

这部分首先将傅里叶变换引入概率密度函数的求解中, 得到特征函数这个很好用的工具, 再借助特征函数推导大数定律和中心极限定理的一般结论, 为数理统计的展开做好铺垫.

大数定律的内容很简单, 就是抽样次数足够大时, 频率近似于概率, 均值近似于数学期望, 这给大样本统计提供了理论依据. 中心极限定理说明多个独立同分布随机变量之和近似于正态分布, 这鼓励我们在大样本统计中使用正态分布进行统计推断.

3.4.1 随机变量的特征函数

1. 定义: $\varphi(t) = E(e^{itX})$. 离散情形下, $\varphi(t) = \sum_{k=1}^{\infty} p_k e^{itX_k}$; 连续情形下, $\varphi(t) = \int_{-\infty}^{\infty} p(x)e^{itx}dx$.

2. 性质: (1) 若 X 与 Y 独立, $Z = X + Y$, 则 $\varphi_Z(t) = \varphi_X(t) \cdot \varphi_Y(t)$.

(2) 求各阶矩的方式: $\varphi^{(k)}(0) = i^k E(X^k)$.

(3) 唯一性定理: 分布函数由特征函数唯一确定.

3. 逆转公式: $F(x_2) - F(x_1) = \lim_{T \rightarrow +\infty} \frac{1}{2\pi} \int_{-T}^T \frac{e^{-itx_1} - e^{-itx_2}}{it} \varphi(t)dt$.

4. 连续随机变量的逆变换公式: $p(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-itx} \varphi(t) dt$.

3.4.2 大数定律

1. 一般形式: 对任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n X_i - \frac{1}{n} \sum_{i=1}^n EX_i\right| < \varepsilon\right) = 1$$

2. 伯努利大数定律: 令 S_n 为 n 重伯努利试验中事件发生的次数, p 为事件发生的概率, 则对任意 $\varepsilon > 0$, 有

$$\lim_{n \rightarrow +\infty} P\left(\left|\frac{S_n}{n} - p\right| < \varepsilon\right) = 1$$

3. 切比雪夫大数定律: 当 $\{X_n\}$ 两两不相关且 $Var(X_i)$ 有界时, 大数定律成立.

4. 辛钦大数定律: 当 X_1, \dots, X_n 独立同分布且 EX_i 存在时, 大数定律成立.

5. 辛钦大数定律的证明: 令 $\varphi(t)$ 为 X_i 共同的特征函数, 数学期望为 a , 将 $\varphi(t)$ 在 $t = 0$ 处泰勒展开: $\varphi(t) = 1 + iat + o(t)$. 故 $\varphi_{\frac{1}{n} \sum_{i=1}^n X_i}(t) = \left[\varphi\left(\frac{t}{n}\right)\right]^n \sim e^{iat}$, 恰是退化分布的特征函数.

3.4.3 中心极限定理

1. 林德伯格-莱维中心极限定理: 令 X_n 独立同分布, $EX_i = \mu$, $DX_i = \sigma^2$, 则 $n \rightarrow \infty$ 时, $\frac{\sum_{i=1}^n X_i - n\mu}{\sigma\sqrt{n}}$ 的分布弱收敛于标准正态分布.

2. 棣莫弗-拉普拉斯中心极限定理: 令 S_n 为 n 重伯努利试验中事件发生的次数, p 为事件发生的概率, 则 $\frac{S_n - np}{\sqrt{np(1-p)}}$ 在 $n \rightarrow \infty$ 时的分布弱收敛于标准正态分布.

3. 中心极限定理的证明: 将 X_i 标准化: $Y_i = \frac{X_i - \mu}{\sigma}$, 则 $EY_i = 0$ 且 $DY_i = 1$. 令 Y_N 的特征函数为 $\varphi(t)$, 故 $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$ 的特征函数为 $\left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n$. 由泰勒展开: $\varphi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o(t^2)$, 在 $n \rightarrow \infty$ 时, $\frac{1}{\sqrt{n}} \sum_{i=1}^n Y_i$ 的特征函数趋近于 $e^{-\frac{t^2}{2}}$, 恰为 $N(0, 1)$ 的特征函数.

3.5 数理统计基本概念

概率论研究的随机变量都有确定的总体, 而现实生活中, 我们通常需要推断某一总体服从何种分布. 这就是数理统计的核心任务: 推断总体服从的分布族, 以及通过样本估计分布中的未知参数. 由于总体的性质只能通过抽样反馈, 因此需要研究由样本推断总体的方法.

在研究过程中, 通常假设各样本与总体同分布且相互独立, 并利用统计量的分布进行推断, 这一思想贯穿了后面的所有章节. 而这部分的任务是打好基础, 理清数理统计的基本概念, 并初步介绍最常用的统计量及其抽样分布, 为参数估计和假设检验打好基础.

3.5.1 统计学基本思想

- 1. 任务:** 收集受随机因素影响的数据, 并根据样本推断总体分布.
- 2. 总体:** 研究对象的全体. 具体分布未知, 一般认为分布族已知, 即推断分布中的未知参数.
- 3. 样本:** 从总体中随机抽取的 n 个数据, 记作 x_1, x_2, \dots, x_n . 若这些样本独立同分布 (i.i.d.), 则称为简单随机样本.
- 4. 统计量:** 当总体分布族已知而参数未知时, 可以构造只与样本有关而与未知参数无关的函数 $T = T(x_1, \dots, x_n)$, 利用统计量的特征估计未知参数.

3.5.2 常用统计量

1. 样本均值: $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$.
2. 样本方差: $s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2$. 计算时常用公式

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^n x_i^2 - n\bar{x}^2 \right)$$

3. 样本标准差: $s = \sqrt{s^2}$.
4. 样本 k 阶矩: $a_k = \frac{1}{n} \sum_{i=1}^n x_i^k$.

3.5.3 抽样分布

1. 定义: 统计量的分布称为抽样分布.
2. χ^2 分布: 若简单随机样本 $x_1, \dots, x_n \sim N(0, 1)$, 则 $\sum_{i=1}^n x_i^2$ 服从自由度为 n 的 χ^2 分布, 记作

$$\sum_{i=1}^n x_i^2 \sim \chi^2(n)$$

其概率密度函数只在第一象限定义且非对称. 不用刻意记忆其具体 p.d.f., 但需注意 $\chi^2(2)$ 的概率密度函数为 $\frac{1}{2}e^{-\frac{1}{2}x}$ ($x > 0$).

3. F 分布: 令独立随机变量 $X \sim \chi^2(m)$, $Y \sim \chi^2(n)$, 则 $\frac{X/m}{Y/n}$ 服从自由度为 m 与 n 的 F 分布, 记作

$$\frac{X/m}{Y/n} \sim F(m, n)$$

F 分布具有的特殊性质: 若 $X \sim F(m, n)$, 则 $\frac{1}{X} \sim F(n, m)$.

4. t 分布: $X \sim N(0, 1)$, $Y \sim \chi^2(n)$, 则 $\frac{X}{\sqrt{Y/n}}$ 服从自由度为 n 的 t 分布, 记作

$$\frac{X}{\sqrt{Y/n}} \sim t(n)$$

t 分布与 F 分布间存在联系: 若 $X \sim t(n)$, 则 $X^2 \sim F(1, n)$.

3.5.4 利用抽样分布统计推断

1. 前置条件: x 服从正态分布, 即 $x \sim N(\mu, \sigma^2)$.

2. σ^2 已知, 对 μ 统计推断: 构造统计量

$$\frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

3. σ^2 未知, 对 μ 统计推断: 由 \bar{x} 与 s^2 相互独立, 可构造统计量

$$\frac{\bar{x} - \mu}{s / \sqrt{n}} \sim t(n - 1)$$

4. 对 σ^2 统计推断:

$$\frac{(n - 1)S^2}{\sigma^2} \sim \chi^2(n - 1)$$

5. 双正态总体方差比推断: 令 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, 构造统计量

$$\frac{s_X^2}{s_Y^2} \sim F(m - 1, n - 1)$$

3.5.5 充分统计量

1. **定义:** 若统计量 T 包含了样本的全部信息, 即给定 T 的取值后, x_1, \dots, x_n 的分布与未知参数 θ 无关, 则称 T 为 θ 的充分统计量.

2. **因子分解定理:** 若总体分布为 $f(x; \theta)$, 存在函数 $g(T, \theta)$ 与 $h(x_1, \dots, x_n)$ 使得 $f(x_1, \dots, x_n; \theta) = g(T, \theta) \cdot h(x_1, \dots, x_n)$, 则 T 为 θ 的充分统计量.

3.6 参数估计

参数估计的目的是对分布族已知, 但含有未知参数的总体, 通过样本估计其中的未知参数. 一种思路为点估计, 即给出参数的确切估计值; 另一种思路为区间估计, 即给出一个大致范围, 有很大的可能性包含参数的真实值. 点估计的评价标准为无偏性和有效性, 即样本越多, 估计值越接近真实值, 且波动尽可能小; 区间估计的手段是通过抽样分布的分位数, 划定统计量所处的范围以包含分布中比例为 $1 - \alpha$ 的部分, 再解出参数所处的范围.

考虑到与机器学习接轨, 这一部分列举了很多超纲的内容, 如用先验推后验的贝叶斯估计, 以及求 ML 估计的 EM 算法, 可视自身需要加以取舍.

3.6.1 矩估计

1. **思想:** 点估计的一种, 另一种即下面讨论的最大似然估计. 令总体分布为 $X(\theta)$. 用样本矩 $a_k = \sum_{i=1}^n X_i^k$ 代替总体矩 $EX^k(\theta)$, 列方程求解未知参数.

2. 以正态分布的矩估计为例: 令 $X \sim N(\mu, \sigma^2)$, 则

$$\begin{cases} EX = \mu &= \frac{1}{n} \sum_{i=1}^n x_i \\ EX^2 = \sigma^2 + \mu^2 &= \frac{1}{n} \sum_{i=1}^n x_i^2 \end{cases}$$

从中解出 $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2 = s_n^2$.

3. 矩估计的相合性: 若 $\lim_{n \rightarrow \infty} E(\hat{\theta}) = \theta$, $\lim_{n \rightarrow \infty} D(\hat{\theta}) = 0$, 则称 $\hat{\theta}$ 为 θ 的相合估计. 矩估计通常为相合估计.

3.6.2 最大似然估计

1. 思想: 选取参数 θ , 使得样本概率 $f(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f(x_i; \theta)$ 最大.
2. 求解方式: 令似然函数 $L(\theta) = \prod_{i=1}^n f(x_i; \theta)$, 求解方程以解出 $\hat{\theta}$:

$$\frac{d \ln L(\theta)}{d\theta} = 0$$

3. 以正态分布的 ML 估计为例: 令 $X \sim N(\mu, \sigma^2)$, 则未知参数 θ 由 μ 和 σ^2 构成. 求解下列方程组:

$$\begin{cases} \frac{\partial \ln L(\mu, \sigma^2)}{\partial \mu} = - \sum_{i=1}^n \frac{\mu - x_i}{\sigma^2} = 0 \\ \frac{\partial \ln L(\mu, \sigma^2)}{\partial \sigma^2} = - \frac{n}{2\sigma^2} + \sum_{i=1}^n \frac{(\mu - x_i)^2}{2\sigma^4} = 0 \end{cases}$$

解得 $\hat{\mu} = \bar{x}$, $\hat{\sigma}^2 = s_n^2$, 恰好与矩估计结果一致.

4. EM 算法: Expectation Maximization, 针对似然函数中存在不可观测的隐

变量 z 时的局部 ML 优化.

(1) E 步: 构造似然函数 $Q(\theta|x, \theta^{(i)}) = E_z[\ln L(\theta; x, z)]$, 目的是消除隐变量 z 的随机性;

(2) M 步: 在已知上轮迭代值 $\theta^{(i)}$ 和样本 x 的情况下, 寻找使似然函数最大的局部最优解:

$$\theta^{(i+1)} = \arg \max_{\theta} Q(\theta|x, \theta^{(i)})$$

(3) 迭代: 设定初始值 $\theta^{(0)}$, 重复 E 步和 M 步直至收敛.

3.6.3 点估计的评价标准

1. 无偏性: 若 $E(\hat{\theta}) = \theta$, 则称 $\hat{\theta}$ 为 θ 的无偏估计.
2. 有效性: 若 $\hat{\theta}_1, \hat{\theta}_2$ 均为 θ 的无偏估计, 且 $D\hat{\theta}_1 > D\hat{\theta}_2$, 则估计 $\hat{\theta}_2$ 比估计 $\hat{\theta}_1$ 更有效.
3. Fisher 信息量: $I(\theta) = E \left[\frac{\partial}{\partial \theta} \ln f(x; \theta) \right]^2$. $I(\theta)$ 越大, 表示总体分布中包含未知参数 θ 的信息越多.
4. Cramer-Rao 不等式: 若 T 为 $g(\theta)$ 的无偏估计, 则

$$DT \geq \frac{[g'(\theta)]^2}{nI(\theta)}$$

若 DT 取到 C-R 下界, 则称 T 为 $g(\theta)$ 的有效估计.

3.6.4 贝叶斯估计

1. 思想: 在抽样之前, 便有关于 θ 的先验信息, 即 θ 服从先验分布 $\pi(\theta)$. 以后验信息

$$\pi(\theta|x) = \frac{f(x|\theta)\pi(\theta)}{\int_{\theta} f(x|\theta)\pi(\theta)d\theta}$$

以后验分布的最大值点作为 θ 的点估计.

2. 朴素贝叶斯分类器: 假设各属性 A_1, \dots, A_n 相互独立, 在得到新的样本 $A_1 = a_1, \dots, A_n = a_n$ 后, 尝试将样本归类: $Y \in y_1, \dots, y_m$.

朴素贝叶斯分类器的最大化目标为: 以样本信息为先验, 寻找可能性最大的分类结果, 即

$$k = \arg \max_k P(Y = y_k | A_1 = a_1, \dots, A_n = a_n)$$

由贝叶斯公式以及朴素假设, 最终优化目标为

$$k = \arg \max_k \prod_{i=1}^n P(A_i = a_i | Y = y_k)$$

每一项都可以通过现有样本点在 $Y = y_k$ 时 $A_i = a_i$ 的占比求出, 选出使优化目标最大的 k , 便可作出最优决策 $Y = y_k$.

3. 共轭先验: 若 $\pi(\theta)$ 与 $\pi(\theta|x)$ 同属一个分布族, 则称该分布族为 θ 的共轭先验分布族, 此时样本的作用仅是将分布族中的未知参数作调整.

3.6.5 区间估计

1. 思想: 区别于点估计, 区间估计的目标是给出 θ 可能的所在区间 $[\hat{\theta}_1, \hat{\theta}_2]$, 使 θ 有 $1 - \alpha$ 的概率落入该区间. 通常使用枢轴量法, 即构造合适的统计量, 利用抽样分布的分位数划定置信限.

2. 单侧区间估计: 若给出 $\hat{\theta}$, 使得 $P(\theta \geq \hat{\theta}) \geq 1 - \alpha$, 则 $\hat{\theta}$ 称为单侧置信下限; 若给出 $\hat{\theta}$, 使得 $P(\theta \leq \hat{\theta}) \geq 1 - \alpha$, 则 $\hat{\theta}$ 称为单侧置信上限. 求解方法与双侧区间估计类似.

3. 单正态分布总体区间估计: 设 $X \sim N(\mu, \sigma^2)$. 主要依据为 1.5.4 节给出的统计量.

(1) σ^2 已知, 对 μ 区间估计: 构造统计量

$$u = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} \sim N(0, 1)$$

则通过 $|u| \leq u_{1-\frac{\alpha}{2}}$ 反解出置信度为 $1 - \alpha$ 时 μ 的置信区间.

(2) σ^2 未知, 对 μ 区间估计: 构造统计量

$$t = \frac{\bar{x} - \mu}{s / \sqrt{n}} \sim t(n-1)$$

则通过 $|t| \leq t_{1-\frac{\alpha}{2}}(n-1)$ 反解出置信度为 $1 - \alpha$ 时 μ 的置信区间.

(3) 对 σ^2 区间估计: 构造统计量

$$\chi^2 = \frac{(n-1)S^2}{\sigma^2} \sim \chi^2(n-1)$$

则通过 $\chi^2 \in [\chi_{\frac{\alpha}{2}}^2(n-1), \chi_{1-\frac{\alpha}{2}}^2(n-1)]$ 反解置信度为 $1 - \alpha$ 时 σ^2 的置信区间.

4. 双独立正态分布总体区间估计: $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$. X 有 m 个样本, 且 Y 有 n 个样本.

(1) σ_1^2 与 σ_2^2 已知, 对 $\mu_1 - \mu_2$ 估计: 构造统计量

$$\frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0, 1)$$

(2) $\sigma_1^2 = \sigma_2^2$ 未知, 对 $\mu_1 - \mu_2$ 估计: 构造统计量

$$\sqrt{\frac{m+n-2}{\frac{1}{m} + \frac{1}{n}}} \frac{\bar{x} - \bar{y} - (\mu_1 - \mu_2)}{\sqrt{(m-1)s_X^2 + (n-1)s_Y^2}} \sim t(m+n-2)$$

(3) 对方差比 $\frac{\sigma_1^2}{\sigma_2^2}$ 估计: 构造统计量

$$\frac{s_X^2/\sigma_1^2}{s_Y^2/\sigma_2^2} \sim F(m-1, n-1)$$

3.7 假设检验

统计学中会有很多假设, 最常见的是假设总体服从正态分布. 但这些假设是不是准确呢? 对这些假设作检验的本质就是反证法: 如果假设是准确的, 则选定统计量应该服从某分布, 但统计量结果在该分布中出现可能性很小, 就推翻假设的正确性.

最常见的假设检验即检验分布的参数是否为某一定值, 如灯泡的寿命是否维持原状 (指数分布的参数是否为某一定值), 是该部分的主要内容. 也有其它的检验目标, 如总体分布的假设是否合理, 样本是否为简单随机样本等, 也将作一定介绍.

3.7.1 基本思想

1. 假设检验的基本问题: 选定原假设 H_0 与备择假设 H_1 , 若 H_0 发生的可能性非常小, 则拒绝原假设而接受备择假设; 若不能拒绝原假设, 则接受原假设. 假设检验的一般问题记作

$$H_0 : \text{_____} \text{ vs } H_1 : \text{_____}$$

2. 检验方法: 先假设 H_0 成立, 结合某一检验统计量的分布给出拒绝域. 若该统计量落入拒绝域, 则拒绝 H_0 , 否则接受 H_0 .

3. 两类错误: 若 H_0 为真, 但统计量落入拒绝域, 则犯了第一类错误 α ; 若 H_0

为假, 但接受了 H_0 , 则犯第二类错误 β .

4. 显著性水平 α : 控制犯第一类错误的可能性 $\leq \alpha$, 即假设 H_0 为真, 检验统计量落入拒绝域的概率应小于等于 α .

3.7.2 正态总体假设检验

1. 单正态总体假设检验: $X \sim N(\mu, \sigma^2)$, 样本量为 n , 构造的统计量依然如 1.5.4 节所述. 由于等式假设和不等式假设仅涉及双侧置信区间和单侧置信区间, 处理手法类似, 故仅以等式假设为例.

(1) σ^2 已知, 检验 $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$: 构造统计量

$$u = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

当 H_0 为真时, $u \sim N(0, 1)$, 即拒绝域为 $\{|u| \geq u_{1-\frac{\alpha}{2}}\}$.

(2) σ^2 未知, 检验 $H_0: \mu = \mu_0$ vs $H_1: \mu \neq \mu_0$: 构造统计量

$$t = \frac{\bar{x} - \mu_0}{s / \sqrt{n}}$$

当 H_0 为真时, $t \sim N(0, 1)$, 即拒绝域为 $\{|t| \geq t_{1-\frac{\alpha}{2}}\}$.

(3) 检验 $H_0: \sigma^2 = \sigma_0^2$ vs $H_1: \sigma^2 \neq \sigma_0^2$: 构造统计量

$$\chi^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

当 H_0 为真时, $\chi^2 \sim \chi^2(n-1)$, 即拒绝域为 $\{\chi^2 \leq \chi_{\frac{\alpha}{2}}^2(n-1), \text{ or } \chi^2 \geq \chi_{1-\frac{\alpha}{2}}^2(n-1)\}$.

2. 双正态总体假设检验: 设 $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$, 且 X 的样本数为 m , Y 的样本数为 n .

(1) σ_1^2 和 σ_2^2 已知, 检验 $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$: 取检验统计量

$$u = \frac{\bar{x} - \bar{y}}{\sqrt{\frac{\sigma_1^2}{m} + \frac{\sigma_2^2}{n}}} \sim N(0, 1)$$

(2) $\sigma_1^2 = \sigma_2^2$ 未知, 检验 $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$: 取检验统计量

$$t = \frac{m+n-2}{\sqrt{\frac{1}{m} + \frac{1}{n}}} \frac{\bar{x} - \bar{y}}{(m-1)s_X^2 + (n-1)s_Y^2} \sim t(m+n-2)$$

(3) $m = n$ 且方差未知的成对样本检验, 检验 $H_0: \mu_1 = \mu_2$ vs $H_1: \mu_1 \neq \mu_2$: 取检验统计量

$$t = \frac{\bar{x} - \bar{y}}{(s_X^2 + s_Y^2)/\sqrt{2n}} \sim t(2n-2)$$

(4) 方差比检验, 检验 $H_0: \sigma_1^2 = \sigma_2^2$ vs $H_1: \sigma_1^2 \neq \sigma_2^2$: 取检验统计量

$$F = \frac{s_X^2}{s_Y^2} \sim F(m-1, n-1)$$

3.7.3 广义似然比检验

1. 思想: 检验 $H_0: \theta \in \Theta$ vs $H_1: \theta \notin \Theta$ 时, 取统计量

$$\Lambda(x_1, \dots, x_n) = \frac{\sup_{\theta \in \Theta} f(x_1, \dots, x_n; \theta)}{\sup_{\theta \notin \Theta} f(x_1, \dots, x_n; \theta)}$$

Λ 越大, 说明 H_0 越有可能成立.

2. 拒绝域: 尚未有统一的形式. 但可以用渐近分布 $2\Lambda \sim \chi^2(n)$, 其中 n 为独立参数的个数.

3.7.4 拟合优度检验

1. 分布拟合检验: 设总体被分为 r 个类 A_1, \dots, A_r , A_i 类中有 n_i 个样本, 检验原假设 $H_0: A_i$ 所占比率为 p_i , 其中 $\sum_{i=1}^r p_i = 1, \sum_{i=1}^r n_i = n$. 构造统计量

$$\chi^2 = \sum_{i=1}^r \frac{(n_i - np_i)^2}{np_i} \sim \chi^2(r-1)$$

拒绝域为 $\{\chi^2 \geq \chi_{1-\alpha}^2(r-1)\}$.

2. χ^2 拟合优度检验: 若 X 的分布函数为 $F(x)$, 将样本归为 r 类: $(-\infty, a_1], (a_1, a_2], \dots, (a_{r-1}, +\infty)$, 每一类理论占比 $p_i = F(a_i) - F(a_{i-1})$, 实际有 n_i 个样本落入第 i 类, 在此基础上应用分布拟合检验.

3.7.5 正态性检验

1. 目的: 检验总体是否服从正态分布.

2. 概率图纸法: 令样本从小到大排列为 $x_{(1)}, \dots, x_{(n)}$, 将点 $(x_{(i)}, \frac{i-0.375}{n+0.25})$ 描在图纸上, 若近似成一条直线, 则认为总体 X 服从正态分布.

3.7.6 游程检验

1. 目的: 检验样本是否随机抽取.

2. 游程检验: 设样本中位数为 M_e , 将样本按抽样时间顺序排列, 并将 $\geq M_e$ 的值替换为 1, $< M_e$ 的值替换为 0, 得到一串 0-1 序列.

3. 判断依据: 把以 0 为界的连续 1 串称为 1 游程, 以 1 为界的连续 0 串称为 0 游程. 若 0 游程数和 1 游程数之和过大或过小, 则拒绝采样随机性, 拒绝域通过查表得出.

3.8 方差分析

单因素方差分析用于解决这一类问题: 控制其它因素都一样, 就改变一个因素, 会不会造成很显著的影响? 换用统计语言来说, 设一个因素有不同的各个水平, 这些水平的均值是否相等? 若相等, 则因素 A 对实验结果没啥影响; 若不相等, 则因素 A 的影响显著. 方差分析作出一个假设: 各水平服从方差相等的正态分布.

样本的波动可由两部分构成: 一是随机性导致同一水平内的数据波动, 即组内误差; 二是因素 A 的作用使不同水平的样本发生了质变, 即组间误差. 显然组间误差占比越高, 因素 A 的影响越显著, 方差分析表也基于此思想得出.

3.8.1 基本思想

1. 检验问题: 设因素 A 有 r 个水平, 各水平均为正态总体 $N(\mu_i, \sigma^2)$ 且方差相等, 检验因素 A 对均值的影响是否显著, 即检验

$$H_0: \mu_0 = \mu_1 = \cdots = \mu_r \quad \text{vs} \quad H_1: \mu_0, \mu_1, \cdots, \mu_r \text{不全相等}$$

2. 统计模型: 令 y_{ij} 表第 i 个总体的第 j 次试验结果, m_i 为水平 A_i 的样本数, 总样本数 $n = \sum_{i=1}^r m_i$. 记 $\varepsilon_{ij} = y_{ij} - \mu_i$ 为随机误差, 则 ε_{ij} 相互独立, 且 $\varepsilon_{ij} \sim N(0, \sigma^2)$.

3. 组内偏差: 令 \bar{y}_i 表示第 i 个总体的组内均值, 则用 $S_e = \sum_{i=1}^r \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2$ 表示第 i 个总体的组内偏差.

4. 组间偏差: 令 \bar{y} 为所有样本的均值, 则用 $S_A = \sum_{i=1}^r m_i (\bar{y}_i - \bar{y})^2$ 表示因素 A 导致的组间偏差.

5. 平方和分解公式: 令总偏差为 $S_T = \sum_{i=1}^r \sum_{j=1}^{m_i} (y_{ij} - \bar{y})^2$, 由 $y_{ij} - \bar{y} = (y_{ij} - \bar{y}_i) + (\bar{y}_i - \bar{y})$ 可推导如下重要公式:

$$S_T = S_A + S_e$$

3.8.2 单因素方差分析

1. 基本思想: 由平方和分解公式, 若 S_A 远大于 S_e , 即偏差大部分由因素 A 导致, 则认为因素 A 影响显著. 构造检验统计量

$$F = \frac{S_A/(r-1)}{S_e/(n-r)} \sim F(r-1, n-r)$$

当 F 大于临界值 $F_{1-\alpha}(r-1, n-r)$ 时, 拒绝原假设, 认为因素 A 显著.

2. 单因素方差分析表

	平方和	自由度	均方	F 比	临界值
因素 A	S_A	$r-1$	$S_A/(r-1)$	$F = \frac{S_A/(r-1)}{S_e/(n-r)}$	$F_{1-\alpha}(r-1, n-r)$
误差 e	S_e	$n-r$	$S_e/(n-r)$		
总和	S_T	$n-1$			

3. 填表方法: 先计算 $S_T = \sum_{i=1}^r \sum_{j=1}^{m_i} (y_{ij} - \bar{y})^2$, 再计算 $S_e = \sum_{i=1}^r \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2$, 由平方和分解公式计算 $S_A = S_T - S_e$, 再从左到右填写剩下内容. 最后比较 F 比和临界值, 若 F 比大于临界值, 则拒绝 H_0 , 认为因素 A 显著.

3.8.3 方差齐性检验

1. 目的: 检验 r 个总体是否符合方差相等的假设条件.

2. 哈特利检验: 令 $H = \frac{\max\{s_1^2, \dots, s_r^2\}}{\min\{s_1^2, \dots, s_r^2\}}$, 则 H 越接近 1, 越有可能认为方差相等. 查表得 H 分布的分位数和拒绝域.

3.9 回归分析

现实生活中, 很难有自变量和因变量的关系能和数学分析中研究的函数一样, 具有良好的性质. 但是我们可以用性质好的函数去拟合变量间的相关关系, 并综合运用前述统计方法, 评价这种拟合到底合不合理, 这就是回归分析的最基本思想.

对单变量关系的情形, 若将样本点 (x, y) 描在图纸上, 仅有散点图很像一条直线时, 我们才能猜测变量间存在线性关系, 其它形状的散点图都不能直接得出结论. 因此一元线性回归是回归分析中最重要的一环, 即用线性函数 $y = \beta_0 + \beta_1 x$ 拟合 x 和 y 间的相关关系.

3.9.1 基本思想

1. 目的: 令 x 为自变量, y 为因变量, 用函数关系 $y = f(x)$ 拟合 x 与 y 间的相关关系, 并要求误差尽可能小.

2. 一元线性回归: 用 $y = \beta_0 + \beta_1 x$ 拟合相关关系, 统计模型为

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \varepsilon_i \sim N(0, \sigma^2)$$

3.9.2 回归系数的最小二乘估计

1. 目的: 令 n 组样本对为 (x_i, y_i) , 求 $\hat{\beta}_0, \hat{\beta}_1$, 使误差和 $Q = \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2$ 最小.

2. 求解: 由 $\frac{\partial Q}{\partial \beta_0} = 0$, 得 $2 \sum_{i=1}^n (\beta_0 + \beta_1 x_i - y_i) = 0$.

由 $\frac{\partial Q}{\partial \beta_1} = 0$, 得 $2 \sum_{i=1}^n x_i (\beta_0 + \beta_1 x_i - y_i) = 0$. 联立解得

$$\hat{\beta}_1 = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2}, \quad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

3. 估计的性质: $\hat{\beta}_0, \hat{\beta}_1$ 均为无偏估计.

3.9.3 区间估计与预测

1. 区间估计目的: 给定 x_0 , 求 $E(y_0) = \beta_0 + \beta_1 x_0$ 的区间估计.

2. 区间估计方法: 令 $\hat{y}_0 = \hat{\beta}_0 + \hat{\beta}_1 x_0$, 构造统计量

$$\frac{\hat{y}_0 - (\beta_0 + \beta_1 x_0)}{\hat{\sigma} \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{l_{xx}}}} \sim t(n-2)$$

其中 $l_{xx} = \sum_{i=1}^n x_i^2 - \frac{1}{n} (\sum_{i=1}^n x_i)^2$.

3. 预测: 给定 x_0 的条件下, 求 y_0 的区间估计. 构造统计量

$$\frac{\hat{y}_0 - y_0}{\hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{l_{xx}}}} \sim t(n-2)$$

从中反解出 y_0 所处的区间.

3.9.4 显著性检验

1. 目的: 检验 y 和 x 的相关性是否显著. 若 y 与 x 无关, 则 $\beta_1 = 0$, 即检验假设 $H_0: \beta_1 = 0$ vs $H_1: \beta_1 \neq 0$.

2. t 检验: 取检验统计量

$$t = \frac{\hat{\beta}_1}{\hat{\sigma} / \sqrt{l_{xx}}} \sim t(n-2)$$

3. F 检验: 取 $S_R = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$, $S_e = \sum_{i=1}^n (y_i - \hat{y}_i)^2$, $S_T = \sum_{i=1}^n (y_i - \bar{y})^2$, 则有平方和分解公式 $S_T = S_R + S_e$. 列方差分析表:

	平方和	自由度	均方	F 比	临界值
回归 R	S_R	1	S_R	$F = \frac{S_R}{S_e/(n-2)}$	$F_{1-\alpha}(1, n-2)$
误差 e	S_e	$n-2$	$S_e/(n-2)$		
总和	S_T	$n-1$			

与方差分析流程一致. 实际上, F 检验与 t 检验等价.

3.9.5 多元线性回归

1. 目的: 用 $y = \omega^T x + b$ 拟合向量 y 与 x 之间的相关关系.
2. 求解: 最小二乘法. 解为 $\theta = (X^T X)^{-1} X^T Y$, 其中 $(x_1, y_1), \dots, (x_m, y_m)$ 为样本点,

$$\theta = \begin{pmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \\ b \end{pmatrix}, \quad X = \begin{pmatrix} x_{11} & x_{12} & \cdots & x_{1n} & 1 \\ x_{21} & x_{22} & \cdots & x_{2n} & 1 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x_{m1} & x_{m2} & \cdots & x_{mn} & 1 \end{pmatrix}, \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

3.9.6 非线性回归

1. 确定函数形式: 根据散点图形状, 确定变换 $z = \varphi(y)$.

-
2. 作变换: 重新绘制 (x, z) 的散点图.
 3. 线性回归: 若 z 与 x 的散点图近似直线, 则对 x 和 z 作一元线性回归.