第 8 期算法班第 1 课微积分线性代数选讲

管老师 @ 七月在线

七月在线

May, 2017

主要内容

- 微积分选讲
  - 极限
  - 微分与泰勒级数
  - 积分与微积分基本定理
  - 牛顿法
  - 参考资料
- 线性代数选讲
  - 线性映射与矩阵
  - 矩阵变换与特征值
  - 奇异值分解
  - 应用举例：PCA
  - 参考资料

极限

通俗语言

函数 f 在 x_0 处的极限为 L

数学记号

$$\lim_{x\to x_0} f(x) = L$$

精确描述:$\epsilon - \delta$ 语言

对于任意的正数 $\epsilon > 0$, 存在正数 δ, 使得任何满足 $|x - x_0| < \delta$ 的 x, 都有

$$|f(x) - L| < \epsilon$$

通俗语言适合于说给对方听，数学记号适合于写给对方看，精确描述比较啰嗦但是非常精确不会造成误解，主要用于证明.

极限: 无穷小阶数

Definition (无穷小阶数)

当 $x \to 0$ 时,

- 如果 $\lim_{x\to 0} f(x) = 0$ 而且 $\lim_{x\to 0} f(x)/x^n = 0$ 那么此时 $f(x)$ 为 n 阶以上无穷小，记为

$$f(x) = o(x^n), x \to 0$$

- 如果 $\lim_{x\to 0} f(x) = 0$ 而且 $\lim_{x\to 0} f(x)/x^n$ 存在且不等于零，那么此时 $f(x)$ 为 n 阶无穷小，记为

$$f(x) = O(x^n), x \to 0$$

为了方便，在不至于引起误解的时候我们回省略掉 $x \to 0$.

所谓无穷小的阶数，就是用我们比较熟悉的多项式类型的无穷小量来衡量其他的无穷小量.

微分学

微分学的核心思想: 逼近.

Definition (函数的导数)

如果一个函数 $f(x)$ 在 x_0 附近有定义，而且存在极限

$$L = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}.$$

那么 $f(x)$ 在 x_0 处可导且导数 $f'(x_0) = L$.

等价定义

无穷小量表述: 线性逼近

如果存在一个实数 L 使得 $f(x)$ 满足,

$$f(x) = f(x_0) + L(x - x_0) + o(x - x_0), x \to x_0.$$

那么 $f(x)$ 在 x_0 处可导且导数 $f'(x_0) = L$.

求导法则

- 链式法则:$\frac{d}{dx}(g \circ f) = \frac{dg}{dx}(f) \cdot \frac{df}{dx}$
- 加法法则:$\frac{d}{dx}(g + f) = \frac{dg}{dx} + \frac{df}{dx}$
- 乘法法则:$\frac{d}{dx}(g \cdot f) = \frac{dg}{dx} \cdot f + g \cdot \frac{df}{dx}$
- 除法法则:$\frac{d}{dx}(\frac{g}{f}) = \frac{\frac{dg}{dx} \cdot f - \frac{df}{dx} \cdot g}{f^2}$
- 反函数求导: $\frac{d}{dx}(f^{-1}) = \frac{1}{\frac{df}{dx}(f^{-1})}$

所有求导法则原则上都可以由链式法则结合二元函数的偏导数来推出来，有兴趣的同学可以思考一下这是为什么

微分学

Definition (函数的高阶导数)

如果函数的导数函数仍然可导，那么导数函数的导数是二阶导数，二阶导数函数的导数是三阶导数. 一般地记为

$$f^{(n)}(x) = \frac{d}{dx} f^{(n-1)}(x)$$

或者进一步

$$f^{(n)}(x) = \frac{d^n}{dx^n} f(x)$$

导数是对函数进行线性逼近，高阶导数是对导数函数的进一步逼近，因为没有更好的办法，所以数学家选择继续使用线性逼近.

用多项式逼近的方式描述高阶导数，我们就得到了泰勒级数.

泰勒/迈克劳林级数: 多项式逼近

如果 $f(x)$ 是一个无限次可导的函数，那么在任何一点 x_0 附近我们可以对 $f(x)$ 做多项式逼近:

$$f(x_0+\Delta_x) = f(x_0) + f'(x_0)\Delta_x + \frac{f''(x_0)}{2}\Delta_x^2 + \cdots + \frac{f^{(n)}(x_0)}{n!}\Delta_x^n + o(\Delta_x^n)$$

在本课中我们不关注对于尾巴上的余项 $o(\Delta_x^n)$ 的大小估计 常庚哲和史济怀老师把泰勒级数称为是一元微分学的顶峰，我也同意这个观点

积分学: 理解积分：无穷求和，体积

Definition (单变量函数黎曼积分)

令 $f(x)$ 为开区间 (a, b) 上的一个连续函数，对于任何一个正整数 n 定义,$x_i = a + \frac{i(b-a)}{n}$ 求和式:

$$S_n(f) = \sum_{i=0}^{n-1} f(x_i)(x_{i+1} - x_i)$$

如果极限 $\lim_{n\to\infty} S_n(f)$ 存在, 那么函数 $f(x)$ 在这个区间上的黎曼积分为

$$\int_a^b f(x)dx = \lim_{n\to\infty} S_n(f)$$

积分学: 理解积分：无穷求和，体积

理解积分

- 代数意义：无穷求和
- 几何意义：函数与 X 轴之间的有向面积

此处课堂画图举例说明

Theorem (牛顿-莱布尼茨公式)

如果 $f(x)$ 是定义在闭区间 $[a,b]$ 上的可微函数, 那么就有

$$\int_a^b f'(t)dt = f(b) - f(a)$$

不定积分表示为

$$\int f'(t)dt = f(x) + C$$

牛顿-莱布尼茨公式展示了微分与积分的基本关系: 在一定程度上微分与积分互为逆运算.

Example

函数 $\ln(x)$ 的不定积分

令 $f(x) = x\ln(x) - x$，则 $f'(x) = 1 \cdot \ln(x) + x \cdot \frac{1}{x} - 1 = \ln(x)$.
根据牛顿-莱布尼茨公式我们得到

$$\int \ln(t)dt = \int f'(t)dt = x\ln(x) - x + C$$

牛顿法

很多机器学习或者统计的算法最后都转化成一个优化的问题. 也就是求某一个损失函数的极小值的问题, 在本课范围内我们考虑可微分的函数极小值问题.

优化问题

对于一个无穷可微的函数 $f(x)$，如何寻找他的极小值点.

极值点条件

- 全局极小值: 如果对于任何 $\tilde{x}$, 都有 $f(x_*) \le f(\tilde{x})$，那么 x_* 就是全局极小值点.
- 局部极小值: 如果存在一个正数 δ 使得，对于任何满足 $|\tilde{x} - x_*| < \delta$ 的 $\tilde{x}$, 都有 $f(x_*) \le f(\tilde{x})$，那么 x_* 就是局部极小值点.（方圆 δ 内的极小值点）
- 不论是全局极小值还是局部极小值一定满足一阶导数/梯度为零，$f' = 0$ 或者 $\nabla f = 0$.

牛顿法

局部极值算法

我们本节课利用极值点条件，来介绍牛顿法.

- 这种方法只能寻找局部极值
- 这种方法要求必须给出一个初始点 x_0
- 数学原理：牛顿法使用二阶逼近
- 牛顿法对局部凸的函数找到极小值，对局部凹的函数找到极大值，对局部不凸不凹的可能会找到鞍点.
- 牛顿法要求估计二阶导数.

牛顿法

牛顿法：二次逼近

首先在初始点 x_0 处，写出二阶泰勒级数

$$
\begin{aligned}
f(x_0+\Delta_x) &= f(x_0) + f'(x_0)\Delta_x + \frac{f''(x_0)}{2}\Delta_x^2 + o(\Delta_x^2) && (1)\\
&= g(\Delta_x) + o(\Delta_x^2) && (2)
\end{aligned}
$$

我们知道关于 Δ_x 的二次函数 $g(\Delta_x)$ 的极值点为 $-\frac{f'(x_0)}{f''(x_0)}$. 那么本着逼近的精神 $f(x)$ 的极值点估计在 $x_0 - \frac{f'(x_0)}{f''(x_0)}$ 附近，于是定义 $x_1 = x_0 - \frac{f'(x_0)}{f''(x_0)}$，并重复此步骤得到序列

$$x_n = x_{n-1} - \frac{f'(x_{n-1})}{f''(x_{n-1})}$$

当初始点选的比较好的时候 $\lim\limits_{n\to\infty} x_n$ 收敛于一个局部极值点.

微积分选讲：参考资料

参考资料

- 数学分析教程，常庚哲，史济怀
- 简明微积分，龚升
- 微积分讲义，陈省身

作业

- 作业，数学分析教程，常庚哲，史济怀 (p142:2,3,7,8; p143:3,4,6; p148:2,3,6; p176:8,11; p210:4,5; p211:6)

线性空间与基

实系数线性空间是一个由向量组成的集合，向量之间可以做加减法，向量与实数之间可以做乘法，而且这些加，减，乘运算要求满足常见的交换律和结合律. 我们也可以类似地定义其他系数的线性空间.

Example (线性空间)

有原点的平面。

- – 如果平面有一个原点 O, 那么平面上任何一个点 P, 都对应着一个向量 $\overrightarrow{OP}$。
- – 这些向量以及他们的运算结构放在一起，就组成一个向量空间。
- 原点 O 在空间中引入了线性结构。(向量之间的加法，以及向量与实数的乘法)

线性空间与基

基是线性空间里的一组向量，使得任何一个向量都可以唯一的表示成这组基的线性组合.

Example (坐标空间)

有原点的平面，加上一组基 $\{\overrightarrow{X}, \overrightarrow{Y}\}$。

- 任何一个向量 $\overrightarrow{OP}$, 都可以唯一表达成 $\overrightarrow{OP} = a\overrightarrow{X} + b\overrightarrow{Y}$ 的形式。
- (a, b) 就是 P 点的坐标。

- 基给出了定量描述线性结构的方法－－坐标系。

线性映射与矩阵

Definition (线性映射)

V 和 W 是两个实线性空间，$T : V \to W$ 如果满足如下条件就是一个线性映射。

$$\begin{aligned} &(i) \quad T(v_1 + v_2) = T(v_1) + T(v_2), && \forall v_1, v_2 \in V \\ &(ii) \quad T(\lambda v) = \lambda T(v), && \forall \lambda \in \mathbb{R}, v \in V \end{aligned}$$

- 线性映射的本质就是保持线性结构的映射
- 到自身的线性映射 $T : V \to V$ 叫做线性变换

线性变换的矩阵描述

V, W 分别为 n, m 维的线性空间，
$\alpha = \{\alpha_1, ..., \alpha_n\}, \beta = \{\beta_1, ..., \beta_m\}$ 分别为 V, W 的一组基。
$T : V \to W$ 是一个线性映射。于是 T, α, β 唯一决定一个矩阵 $A_{\alpha,\beta}(T) = [A_{ij}]_{m\times n}$，使得

$$T(\alpha_j) = \sum_{i=1}^{m} A_{ij} * \beta_i, \forall j \in 1, ..., n \tag{3}$$

(3) 等价于

$$T(\alpha_1, ..., \alpha_n) = (\beta_1, ..., \beta_m) \cdot A_{\alpha,\beta}(T) \tag{4}$$

简记为

$$T(\alpha) = \beta \cdot A_{\alpha,\beta}(T) \tag{5}$$

线性映射与矩阵

如果我们选取 V, W 的另外一组基，$\tilde{\alpha} = \alpha \cdot P$, $\tilde{\beta} = \beta \cdot Q$. 那么存在矩阵 $A_{\tilde{\alpha},\tilde{\beta}}(T)$ 使得,

$$T(\tilde{\alpha}) = \tilde{\beta} \cdot A_{\tilde{\alpha},\tilde{\beta}}(T)$$

两边分别代入 $\tilde{\alpha}$ 与 $\tilde{\beta}$ 得到,

$$T(\alpha) \cdot P = T(\alpha \cdot P) = \beta \cdot Q \cdot A_{\tilde{\alpha},\tilde{\beta}}(T)$$

与(5)比较我们得到矩阵变换公式:

$$Q \cdot A_{\tilde{\alpha},\tilde{\beta}}(T) \cdot P^{-1} = A_{\alpha,\beta}(T) \tag{6}$$

线性映射与矩阵

小结 (线性映射与矩阵)

- 矩阵是线性映射在特定基下的一种定量描述

$$T(\alpha) = \beta \cdot A_{\alpha,\beta}(T)$$

- 基变换下的矩阵变换公式的推导方法

$$Q \cdot A_{\tilde{\alpha},\tilde{\beta}}(T) \cdot P^{-1} = A_{\alpha,\beta}(T)$$

相似变换 (把矩阵看做线性映射)

如果 $T: V \to V$ 是一个线性变换, 那么对于 V 的两组基 α 与 $\tilde{\alpha} = \alpha \cdot P$，线性变换 T 的矩阵分别为

$$A_\alpha(T) \text{ and } A_{\tilde{\alpha}}(T) = P^{-1} \cdot A_\alpha(T) \cdot P$$

方阵的相似变换

- 如果两个方阵 A 和 $\tilde{A}$ 满足，$\tilde{A} = P^{-1}AP$. 那么这两个方阵就互为相似矩阵
- 相似矩阵的几何意义是同一个线性变换在不同的基下的表达形式
- 当研究对象是线性变换的时候，我们只关心矩阵在相似变换下不变的几何性质。

相似变换 (把矩阵看做线性映射)

相似变换下不变的性质

- 行列式 (det)

$$\begin{aligned}\det(P^{-1}AP) &= \det(P^{-1})\det(A)\det(P)\\ &= \det(P^{-1})\det(P)\det(A)\\ &= \det(A)\end{aligned}$$

- 迹 (trace),$\mathrm{tr}(AB) = \mathrm{tr}(BA)$

$$\mathrm{tr}(P^{-1}AP) = \mathrm{tr}(APP^{-1}) = \mathrm{tr}(A \cdot I) = \mathrm{tr}(A)$$

- 秩 (rank)

相似不变量

相似变换下不变的性质

- 特征值：特征方程 $\det(A-\lambda I)=0$ 的根。
  如果 $\det(A-\lambda I)=0$，那么 $\det(P^{-1}(A-\lambda I)P)=0$，于是 $\det(P^{-1}AP-\lambda I)=0$

- 特征值是最重要的相似不变量，利用这个相似不变量可以方便的得出上面所有的不变量。

相合变换 (二次型)

假设 V 是一个实系数线性空间，那么线性空间上的度量指的是空间中向量的内积关系 $G(v_1, v_2)$. 如果 $\alpha\{\alpha_1, \cdots, \alpha_k\}$ 是空间 V 的一组基，那么这个内积一般可以用一个对称矩阵
$H_\alpha = [h_{ij}]_{n\times n}$ 来表示.

$$h_{ij} = G(\alpha_i, \alpha_j)$$

这时候对于任意两个向量 v_1, v_2, 如果 $v_1 = \alpha \cdot x_1, v_2 = \alpha \cdot x_2$，那么

$$G(v_1, v_2) = x_1^T H_\alpha x_2$$

相合变换 (二次型)

方阵的相合变换

- 如果两个对称方阵 A 和 $\tilde{A}$ 满足，$\tilde{A} = P^T AP$. 那么这两个方阵就互为相合矩阵
- 相似矩阵的几何意义是同一个内积结构在不同基下的表示形式

相合变换 (二次型)

方阵的正交相似变换

正交相似变换同时满足相似与相合变换的条件，也就是说它同时保持了矩阵的相似与相合不变量。

- 如果两个对称方阵 A 和 $\tilde{A}$ 满足，$\tilde{A} = P^T AP$. 而且 P 是正交矩阵:$P^T = P^{-1}$. 那么这 A 与 $\tilde{A}$ 就互为正交相似.

相合变换 (二次型)

方阵的正交相似标准型

任何一个对称矩阵 A 都可以正交相似于一个对角矩阵 D.

总存在一个正交矩阵 P 使得，$A = P^T DP$.

相合变换 (二次型)

长方矩阵的奇异值分解 (SVD)

对于任何一个矩阵 $B_{m\times n}$, 存在正交矩阵 $P_{m\times m}$,$Q_{n\times n}$. 使得

$$B = PDQ$$

其中 $D_{m\times n}$ 是一个只有对角元素不为零的矩阵.

应用举例：PCA

主成分分析 (PCA)

PCA 的主要目的是降维，也可以起到分类的作用

- 当数据维度很大的时候，如果相信大部分变量之间存在线性关系，那么我们就希望降低维数，用较少的变量来抓住大部分的信息.
- 一般来讲做 PCA 之前要做 normalization 使得变量中心为 0，而且方差为 1.

比较广泛应用于图像识别，文档处理，推荐系统

应用举例：PCA

主成分分析 (PCA)

- 首先计算变量之间的协方差矩阵 Σ(利用样本)
- 找到 Σ 的正交相似标准型

正交相似标准性的求解由计算机完成，我们主要关心他的几何意义

应用举例：PCA

推荐系统

如果一个旅游网站里面有 10000000 个注册用户，以及 40000 个注册酒店. 网站有用户通过本网站点击酒店页面的记录信息.$A=[A_{ij}]_{10000000\times 40000}$,$A_{ij}$ 表示第 i 个用户点击 j 酒店的次数.

- 如何评价酒店之间的相似度?
- 给定一个酒店，请找出与它最相似的其他几个酒店?
- 如果要给酒店分类，有什么办法?

线性代数：参考资料

参考资料

- 陶哲轩的讲义
  http://www.math.ucla.edu/~tao/resource/general/115a.3.02f/
- 入门教材：线性代数及其应用, 莱 (Lay D.C.) (作者), 刘深泉 (译者)
- 艰深教材：线性代数，李炯生、查建国

作业

- 入门教材：线性代数及其应用, 莱 (Lay D.C.) (作者), 刘深泉 (译者)
  p69:8,13,15,16; p79:7,14
  p185:18,30; p186:9,14,15
  p242:6,17,18; p292:5,6
  p293:19,20,22,25,26; p99:25,26,27,35
  p413:9,11; p423:17,18,21
  p431:1

谢谢大家!