极客大学

# 极客大学机器学习训练营
# 机器学习数学基础

王然

众微科技 AI Lab 负责人

二○二一年五月五日

# 大纲


1 大纲

2 微积分

3 矩阵论和线性代数

4 补充材料

5 参考文献

# 大纲

1 大纲

2 微积分

3 矩阵论和线性代数

4 补充材料

5 参考文献

# 综述



- 在本章中，我们主要复习**微积分**和**矩阵代数**的知识；
- 我们的目的是尽可能通过**非严谨的数学推导**对已有知识进行复习；
- 除去对内容的讲解，我们还会增加**习题**和**补充内容**。这些内容不是必须要理解的，但是是对学习数学有很大帮助的；
- **请注意：任何关于数学的练习都可以极大的提升数学能力**

# 大纲

极客大学

1 大纲

2 微积分
- 导数
- 导数的应用
- 积分

3 矩阵论和线性代数

4 补充材料

5 参考文献

# 主要内容



- 极限；
- 导数；
- 积分。

# 极限的官方定义

极客大学

- 我们说 $x_n$ 收敛于 $x$，并用 $x_n \to x$ 表示，当且仅当对于任何 $\epsilon > 0$，存在 $N$, 使得所有 $n > N$ 的，我们均有 $\|x_n - x\| < \epsilon$。
- 换句话说，当 $n$ 足够大的时候，$x_n$ 可以无限接近 $x$。
- 我们一般也写作 $\lim_{n\to\infty} x_n = x$；通常我们会省略下标
- 一般来说，我们不会直接验证该定义。

# 连续函数

- 我们说 $f$ 是一个连续函数，当且仅当对于任何 $x_n \to x$，$f(x_n) \to f(x)$。
- 我们接触的大部分（?）函数都是连续函数。

# 大纲




1 大纲

2 微积分
- 导数
- 导数的应用
- 积分

3 矩阵论和线性代数

4 补充材料

5 参考文献

# 导数

- 在一定的条件下，给定函数 $f: x \mapsto f(x)$，我们定义 $\frac{df}{dx}|_{x_0} = \lim_{x \to x_0} \frac{f(x)-f(x_0)}{x-x_0}$；
- 导数的含义是在局部的切线；
- 连续的函数**不一定**可导；
- 在没有异议的情况下，我们也可以用 $f'$ 来表示函数 $f$ 的导数。

# 导数的性质和计算



- 大部分的导数都有常见的计算法则；
- 一些常见函数：
  - $\frac{de^x}{dx} = e^x$;
  - $\frac{d\log x}{dx} = \frac{1}{x}$;
  - $\frac{dx^{alpha}}{dx} = \alpha x^{\alpha-1}$;

# 导数的四则运算



- $\frac{d(f(x) \pm \alpha g(x))}{dx} = \frac{d(f(x)}{dx} \pm \alpha \frac{dg(x)}{dx}$；
- $\frac{d(f(x)g(x))}{dx} = \frac{df(x)}{dx} g(x) + \frac{dg(x)}{dx} f(x)$;
- $\frac{df(x)/g(x)}{dx} = \frac{(df(x)/dx)g(x) - (dg(x)/dx)f(x)}{g^2(x)}$；

问题： 可否从第二个式子推导出第三个式子？

# 锁链求道法则



$$\frac{d(f \circ g(x))}{dx} = \frac{df(g)}{dg}\frac{dg(x)}{dx}$$

# 求导练习

$$f : x \mapsto e^{e^x}$$

# 求导练习



$$f: x \mapsto \frac{1}{1+e^{-x}}$$

$$f: x \mapsto x^x$$

# 关于导数的说明

- 对于绝大多数函数而言，导数均是可以求得的；
- 不愿意手算可以采用Scipy。

# 偏导数



- 函数中可能有多个输入，例如 $f: x, y \mapsto x^2 + e^y$;
- 在这种情况下，如果我们只对其中一个函数的导数感兴趣（偏导数）我们一般写作 $\frac{\partial f}{\partial x} = 2x$;
- 复合函数求导仍然有类似的锁链法则，我们将会在第四章矩阵部分讲解。

# 大纲




1 大纲

2 微积分

- 导数
- 导数的应用
- 积分

3 矩阵论和线性代数

4 补充材料

5 参考文献

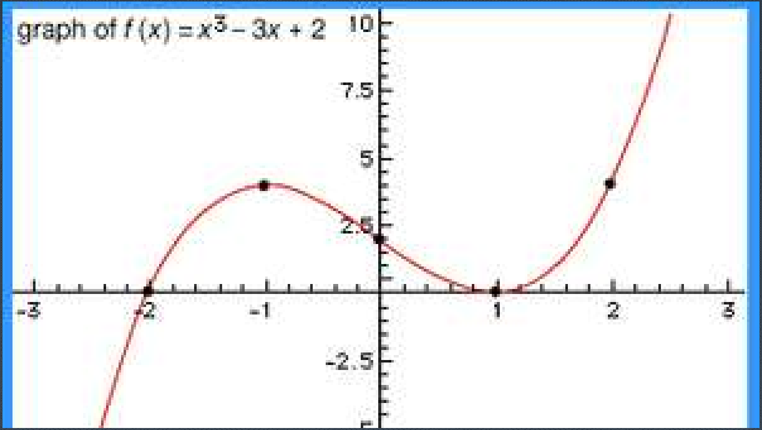
# 无约束求解



- 对于连续可导的函数，一般来说找到（局部）最大/最小值的方式是找到导数为 0 的点；
- 请注意导数为 0 并不意味着这一点是最大/最小值；

# 导数和优化





# 有约束的优化



- 目标为 $\max_{x_1,\cdots,x_n} f(x_1,\cdots,x_n)$;
- 约束为 $g_j(x_1,\cdots,x_n) \geq 0, j = 1,\cdots,J$ 和 $h_k(x_1,\cdots,x_n) = 0, k = 1,\cdots,K$;
- 定义拉格朗日乘子为 $L(\mathbf{x},\lambda,\mu) = f(\mathbf{x}) + \lambda \cdot \mathbf{g}(\mathbf{x}) + \mu \cdot \mathbf{h}(\mathbf{x})$。

# KKT 条件



在一定的条件下，最优解 $\mathbf{x}^*$ 满足：

- $\frac{\partial L}{\partial \mathbf{x}}\left(\mathbf{x}^*, \lambda^*, \mu^*\right) = \mathbf{0}$;
- $\frac{\partial L}{\partial \mu}\left(\mathbf{x}^*, \lambda^*, \mu^*\right) = \mathbf{h}\left(\mathbf{x}^*\right) = \mathbf{0}$;
- $\frac{\partial L}{\partial \lambda}\left(\mathbf{x}^*, \lambda^*, \mu^*\right) = \mathbf{g}\left(\mathbf{x}^*\right) \geq \mathbf{0}, \lambda^* \geq \mathbf{0}$;
- $\lambda^* \odot \frac{\partial L}{\partial \lambda}\left(\mathbf{x}^*, \lambda^*, \mu^*\right) = \lambda^* \odot \mathbf{g}\left(\mathbf{x}^*\right) = \mathbf{0}$。

由于我们在课程中只有少数应用（例如 SVM）使用该方法，所以我们省去练习题。

# 大纲

极客大学

1 大纲

2 微积分
- 导数
- 导数的应用
- 积分

3 矩阵论和线性代数

4 补充材料

5 参考文献

# 牛顿-莱布尼茨公式

极客大学

- 假设 $F$ 的导数是 $f$，并且满足各种良好的性质，我们有 $\int_a^b f(x)dx = F(b) - F(a)$；
- 类似的，我们如果有 $F(x) = \int_a^x f(t)dt + C$，则 $F$ 的导数是 $f$;
- 积分是线性泛函，换句话说 $\int (f(x) + \lambda g(x))\, dx = \int f(x)dx + \lambda \int g(x)dx$。

# 积分的运算

极客大学

- 与导数的运算不同，大部分时候积分是无法得到解析解的；
- 在本课程中，我们仅仅会在有限的部分中应用积分的一些法则。

# 积分的法则：换元法



假设 $\phi(\cdot)$ 是一个单调函数

$$\int_a^b F(\phi(x))\phi'(x)dx = \int_{\phi(a)}^{\phi(b)} F(y)dy$$

# 换元法练习



求解下列的积分：

$$\int_0^1 x\sqrt{1+x^2}dx$$

# 分部积分法

极客大学

$$\int_a^b u(x)v'(x)dx = u(x)v(x)|_a^b - \int_a^b u'(x)v(x)dx$$

# 分部积分法练习



求解下列积分：

$$\int_0^1 x^2 e^x dx$$

# 一个特殊的积分



- 令 $\phi_{\mu,\sigma}(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$，则我们可以证明 $\int_{-\infty}^{\infty} \phi_{\mu,\sigma}(x)dx = 1$；
- 该积分在概率论和统计学当中有非常重要的性质；
- （思考题）求解 $\int_{-\infty}^{\infty} e^{\lambda x}\phi_{0,1}(x)dx$。

# 多重积分

- 由于时间关系，我们省略关于多重积分的内容；
- 对于大部分情况而言，多重积分的处理和分别对不同积分元进行处理相差不多。

# 大纲

极客大学

1. 大纲
2. 微积分
3. 矩阵论和线性代数
   - 矩阵的定义和运算
   - 线性代数
4. 补充材料
5. 参考文献

# 大纲

极客大学

1 大纲

2 微积分

3 矩阵论和线性代数

- 矩阵的定义和运算
- 线性代数

4 补充材料

5 参考文献

# 向量的定义



- 在本课程当中，所有的向量均为**列向量**！
- 我们一般表示为（例子中为三维向量）$x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$
- 为了节省空间，我们通常写作 $x = [x_1 \ x_2 \ x_3]^t$，其中上标 $t$ 表示转置。

# 矩阵



- 与向量类似，我们一般将矩阵表示为（以下为 $3 \times 4$ 矩阵）

$$X = \begin{bmatrix} x_{11} & x_{12} & x_{13} & x_{14} \\ x_{21} & x_{22} & x_{23} & x_{24} \\ x_{31} & x_{32} & x_{33} & x_{34} \end{bmatrix}$$

- 对于矩阵的转置，设 $Y = X^t$，则 $Y_{ij} = X_{ji}$；
- 请注意：在大部分数学教材当中，向量/矩阵并不会用特殊的符号进行表示！

# 向量/矩阵的加减法

- 向量和矩阵的加减法只有当二者都是维度相等时候才可以进行；
- 对于向量来说，两者的列数相等；
- 对于矩阵来说，如果 $X$ 为 $M \times N$ 的矩阵，则 $Y$ 也必须是 $M \times N$ 的矩阵才能进行相加减；

# 向量/矩阵的乘法

- 设 $X$ 为 $M \times N$ 的矩阵，$y$ 为 $N$ 维向量，则 $Z = Xy$ 为 $M$ 维向量，且 $Z_i = \sum_j X_{ij} y_j$；
- 设 $X$ 为 $M \times N$ 的矩阵，$Y$ 为 $N \times P$ 的矩阵，则 $Z = XY$ 为 $M \times P$ 的矩阵，并且 $Z_{ik} = \sum_j X_{ij} Y_{jk}$。

# 矩阵乘法练习



- $X = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}$
- $Y = \begin{bmatrix} 7 & 8 \\ 2 & 9 \end{bmatrix}$
- 求 $XY$;

# 矩阵乘法的一些性质

极客大学

- 在大部分时候，矩阵乘法的性质和标量的性质是一样的；
- $X+Y=Y+X$;
- $(X+Y)Z=XZ+YZ$ ; $Z(X+Y)=ZX+ZY$;
- $XYZ=(XY)Z=X(YZ)$ ;
- 但是矩阵的乘法不满足**交换律**，即 $XY\neq YX$;
- 最后对于转置，我们有 $(XY)^t=Y^tX^t$ 以及 $(X+Y)^t=X^t+Y^t$。

# 关于分块矩阵的一些说明

- 一种常用的方法是讲矩阵中的行或列写成向量的形式；
- 例如 $X = \begin{bmatrix} 1 & 3 \\ 2 & 4 \\ 3 & 6 \end{bmatrix}$，我们可以让 $x_1 = [1\ 3]^t$，则 $X = [x_1^t\ x_2^t\ x_3^t]^t$（请自己验证）；
- 将矩阵写成分块形式后，我们仍然“假设”其得到的结果满足矩阵相乘的规律。

# 练习



- 设 $X$ 为 $M \times N$ 的矩阵，$Y$ 为 $N \times P$ 的矩阵，$Z = XY$;
- 请用求和和分块矩阵的方式写出 $Z_{ik}$ 的元素。

# 矩阵的内积和点积

- 假设 $X, Y$ 均为 $M \times N$ 的矩阵；
- $X \odot Y$ 表示对应元素的乘积，其结果为 $M \times N$ 的矩阵；
- $X \cdot Y$ 表示对应元素的乘积的求和，其结果为标量。
- 对于向量来说，以上定义也成立；
- 注意：对于内积来说，有时我们也用 $\langle \mathbf{u}, \mathbf{v} \rangle$ 表示内积

# 大纲

极客大学

1 大纲

2 微积分

3 矩阵论和线性代数
- 矩阵的定义和运算
- 线性代数

4 补充材料

5 参考文献

# 线性空间

- 给定向量 $\{x_1, x_2, \ldots, x_k\}$ 以及实数 $\{\lambda_1, \lambda_2, \ldots, \lambda_k\}$；
- 当 $\lambda_k \in R$ 进行取值时，$\sum_k \lambda_k x_k$ 所构成的为线性空间；
- 在这里，我们可以将 $x_k$ 想想成为坐标轴，而 $\lambda_k$ 则为坐标取值。

# 施密特正交化

极客大学

- 对于上面提到的向量而言，为了充分表达其作用，我们希望向量之间是正交的；换句话说，两者的内积等于 0；
- 给定任何一组向量，我们可以采用施密特正交化得到一组正交的基；
- 施密特正交化见施密特正交化文档

# 大纲



1 大纲
2 微积分
3 矩阵论和线性代数
4 补充材料
5 参考文献

- 大部分的微积分教程均涵盖了我们讲述的内容，一个例子可见北大的数学分析 A；
- 一些教材当中提及了微分的问题，如果想要深入理解微分的概念，请参考Lang (2012)；
- 目前来说，最为完整的矩阵论教材为Schott (2016)。

# 大纲

1 大纲

2 微积分

3 矩阵论和线性代数

4 补充材料

5 参考文献

Lang, Serge (2012). *Fundamentals of differential geometry*. Vol. 191. Springer Science & Business Media.

Schott, James R (2016). *Matrix analysis for statistics*. John Wiley & Sons.

# Thanks!