5 月深度学习班：机器学习中的数学

程博士

julyedu.com

1. 数学理论是算法的基石，编程是工具。三者都很重要，但是如果只会重复编程，不可能理解到事物的本质。
2. 微积分、概率、线性代数和矩阵是优化的基础，优化贯穿几乎所有的工科，人们总是希望求得最优解。机器学习中大量的问题最终都可以归结为一个优化问题(例如SVM)。
3. 2个小时回顾四门极其重要的数学课，选取精华中的精华部分。

本次课件主要参考资料

1. 本人矩阵理论学习笔记
2. 张贤达，矩阵分析与应用
3. 本人凸优化理论学习笔记
4. Stephen Boyd, Convex Optimization, 英文原版
5. 概率和数理统计，本科教材
6. 互联网相关搜索资料

示例表

红色框表示非常重要的定理或内容

定理或内容，仔细弄明白

绿色框表示具体的例子

举例，会举一反三

微积分总视图(板书)

导数(标量)

- 导数的定义

$$f'(a) = \lim_{h \to 0} \frac{f(a+h) - f(a)}{h} \tag{1}$$

- 常见函数的导数
  - $(x^a)' = ax^{a-1}$
  - $(e^x)' = e^x$
  - $(a^x)' = \ln(a)\, a^x$
  - $(\ln(x))' = \frac{1}{x}$
  - $\frac{d}{dx}\sin(x) = \cos(x)$
  - $\frac{d}{dx}\cos(x) = -\sin(x)$

导数法则

- $(\alpha f+\beta g)'=\alpha f'+\beta g'$
- $(fg)'=f'g+fg'$
- $\left(\frac{f}{g}\right)'=\frac{f'g-fg'}{g^2}$
- 如果，$f(x)=h(g(x))$，则有$f'(x)=h'(g(x))\cdot g'(x)$

计算$f(x)=x^4+\sin(x^2)-\ln(x)e^x+7$的导数

$$\begin{aligned} f'(x) &= 4x^{(4-1)}+\frac{d\left(x^2\right)}{dx}\cos(x^2)-\frac{d\left(\ln x\right)}{dx}e^x-\ln(x)\frac{d\left(e^x\right)}{dx}+0 \\ &= 4x^3+2x\cos(x^2)-\frac{1}{x}e^x-\ln(x)e^x \end{aligned}$$

梯度和Hessian矩阵

梯度和Hessian矩阵(以下均假设连续可导)

- 一阶导数和梯度(gradient vector)

$$f'(x); \qquad \nabla f(\mathbf{x}) = \frac{\partial f(\mathbf{x})}{\partial \mathbf{x}} = \begin{bmatrix} \frac{\partial f(\mathbf{x})}{\partial x_1} \\ \vdots \\ \frac{\partial f(\mathbf{x})}{\partial x_n} \end{bmatrix} \tag{2}$$

- 二阶导数和Hessian矩阵

$$f''(x); \qquad \mathbf{H}(\mathbf{x}) = \nabla^2 f(\mathbf{x}) = \begin{bmatrix} \frac{\partial^2 f(\mathbf{x})}{\partial x_1^2} & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_2} & \cdots & \frac{\partial^2 f(\mathbf{x})}{\partial x_1 \partial x_n} \cdots \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_2 \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_2^2} & & \\ & & \ddots & \\ \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_1} & \frac{\partial^2 f(\mathbf{x})}{\partial x_n \partial x_2} & & \frac{\partial^2 f(\mathbf{x})}{\partial x_n^2} \end{bmatrix} \tag{3}$$

二次型的梯度(详见张矩阵)

若 $\boldsymbol{f}(\boldsymbol{x})=[x_1,x_2,\cdots,x_n]$，则

$$\boxed{\frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \boldsymbol{x}}=\boldsymbol{I}} \tag{5.1.15}$$

式中，$\boldsymbol{I}$ 为单位矩阵。这是一个非常有用的结果。

例 5.1.1 若 $\boldsymbol{A}$ 和 $\boldsymbol{y}$ 均与向量 $\boldsymbol{x}$ 无关，则

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{y}}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{x}^{\mathrm{T}}}{\partial \boldsymbol{x}}\boldsymbol{A}\boldsymbol{y}=\boldsymbol{A}\boldsymbol{y}$$

例 5.1.2 注意到 $\boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\langle \boldsymbol{A}^{\mathrm{T}}\boldsymbol{y},\boldsymbol{x}\rangle=\langle \boldsymbol{x},\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}\rangle=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$，故

$$\frac{\partial \boldsymbol{y}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}}{\partial \boldsymbol{x}}=\boldsymbol{A}^{\mathrm{T}}\boldsymbol{y}$$

例 5.1.3 由于

$$\boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}=\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}x_ix_j$$

可求出梯度 $\dfrac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}$ 的第 k 个分量为

$$\left[\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}\right]_k=\frac{\partial}{\partial x_k}\sum_{i=1}^{n}\sum_{j=1}^{n}A_{ij}x_ix_j=\sum_{i=1}^{n}A_{ik}x_i+\sum_{j=1}^{n}A_{kj}x_j$$

即有

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=\boldsymbol{A}\boldsymbol{x}+\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x}$$

特别地，若 $\boldsymbol{A}$ 为对称矩阵，则

$$\frac{\partial \boldsymbol{x}^{\mathrm{T}}\boldsymbol{A}\boldsymbol{x}}{\partial \boldsymbol{x}}=2\boldsymbol{A}\boldsymbol{x}$$

泰勒级数与极值

泰勒级数展开(标量)

- 输入为标量的泰勒级数展开

$$f(x_k+\delta) \approx f(x_k) + f'(x_k)\,\delta + \frac{1}{2} f''(x_k)\,\delta^2$$

- 称满足$f'(x_k)=0$的点为平稳点(候选点)，此时如果还有：
  - $f''(x_k)>0$，x_k为一严格局部极小点(反之，严格局部最大点)(充分条件)
  - 如果$f''(x_k)=0$，有可能是一个鞍点(saddle point)，why?

- 思考实际使用中的局限?

泰勒级数与极值

泰勒级数展开(矢量)(和标量情况对比)

- 输入为矢量的泰勒级数展开

$$f(\mathbf{x}_k+\boldsymbol{\delta}) \approx f(\mathbf{x}_k)+\nabla^T f(\mathbf{x}_k)\,\boldsymbol{\delta}+\frac{1}{2}\boldsymbol{\delta}^T\nabla^2 f(\mathbf{x}_k)\,\boldsymbol{\delta}$$

- 称满足$\nabla^T f(\mathbf{x}_k)=0$的点为平稳点(候选点)，此时如果还有：
  - $\nabla^2 f(\mathbf{x}_k)\succ 0$，$x_k$为一严格局部极小点(反之，严格局部最大点)
  - 如果$\nabla^2 f(\mathbf{x}_k)$不定矩阵，是一个鞍点(saddle point)
  - 思考$\nabla^2 f(\mathbf{x}_k)\succeq 0$

- 梯度方向？梯度下降法从哪儿来？

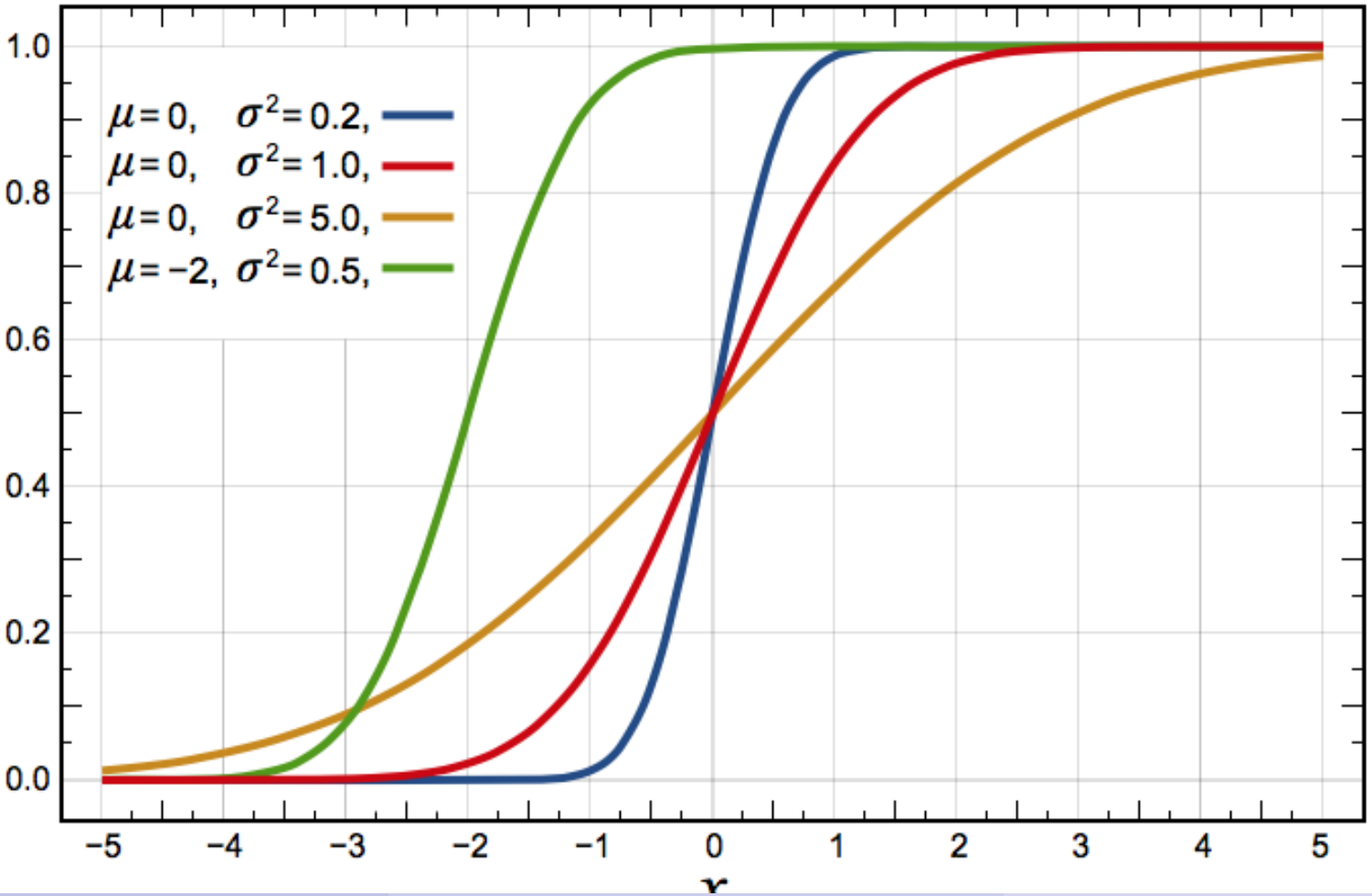
微积分总结(板书)

随机变量(随机事件的数量表现)

- 累积分布函数

$$F_X(x) = \mathrm{P}(X \leq x) \tag{4}$$

$$\mathrm{P}(a < X < b) = F_X(b) - F_X(a) \tag{5}$$



随机变量

- 概率密度函数

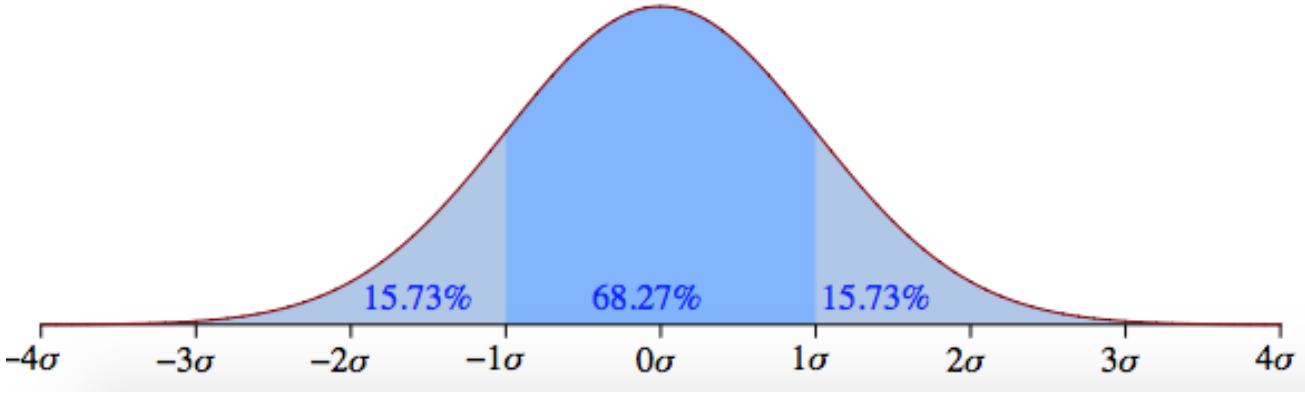
$$f_X(x) = \frac{d}{dx} F_X(x) \tag{6}$$

$$\mathrm{P}[a \leq X \leq b] = \int_a^b f_X(x)\,dx \tag{7}$$

$$F_X(x) = \int_{-\infty}^{x} f_X(u)\,du \tag{8}$$

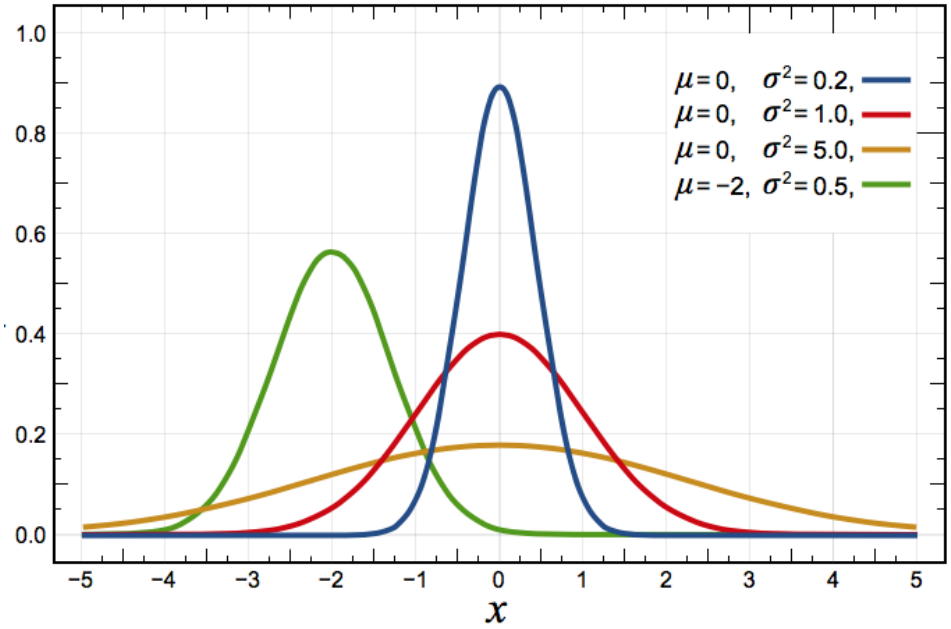
$$\mathrm{P}(t < X < t + dt) = f(t)\,dt$$

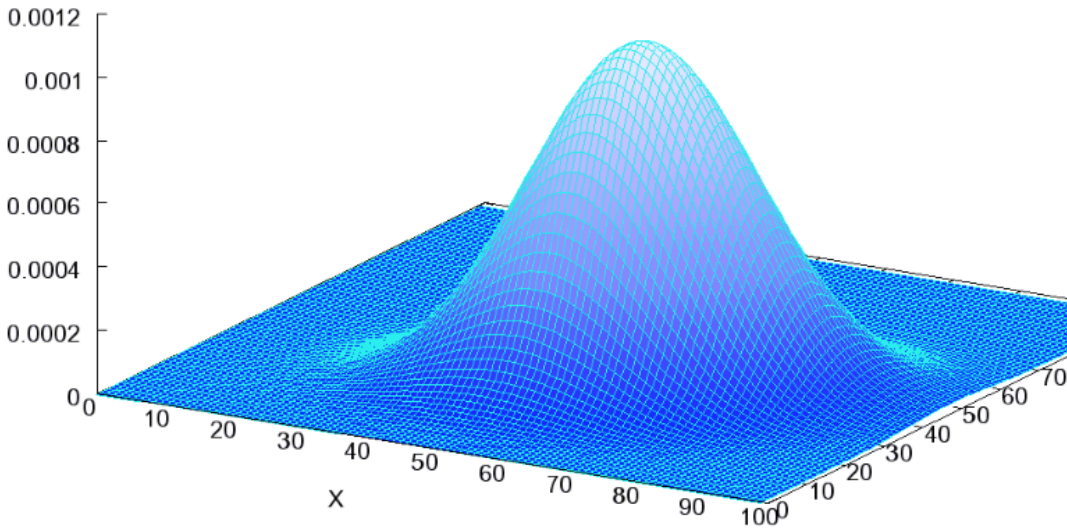
- 举例



高斯分布(最美的分布)

- 一元概率密度: $f(x|\mu,\sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$
- 多元概率密度: $f_{\mathbf{x}}(x_1,\ldots,x_k) = \frac{1}{\sqrt{(2\pi)^k|\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^{\mathrm{T}}\boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right)$
- 中心极限定理
- 思考独立高斯变量相加？思考高斯变量加任意一个随机变量？





贝叶斯公式(机器学习中最重要的公式)

- 通常，$P(A|B) \neq P(B|A)$，但是如何确定两者的关系？（溯源）
  - $P(A|B) = \frac{P(A \cap B)}{P(B)}$，同样有$P(B|A) = \frac{P(A \cap B)}{P(A)}$，因此有$P(A|B)P(B) = P(A \cap B) = P(B|A)P(A)$
  - 可得$P(B|A) = \frac{P(A|B)P(B)}{P(A)}$和$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$
  - 先验概率，后验概率解释
- 二则一：$P(B) = P(A,B) + P(A^C,B) = P(B|A)P(A) + P(B|A^C)P(A^C)$故有

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^C)P(A^C)} \tag{9}$$

- 概率密度形式：

$$f(x|y) = \frac{f(x,y)}{f(y)} = \frac{f(y|x)f(x)}{f(y)} = \frac{f(y|x)f(x)}{\int_{-\infty}^{\infty} f(y|x)f(x)dx} \tag{10}$$

贝叶斯公式

贝叶斯应用举例

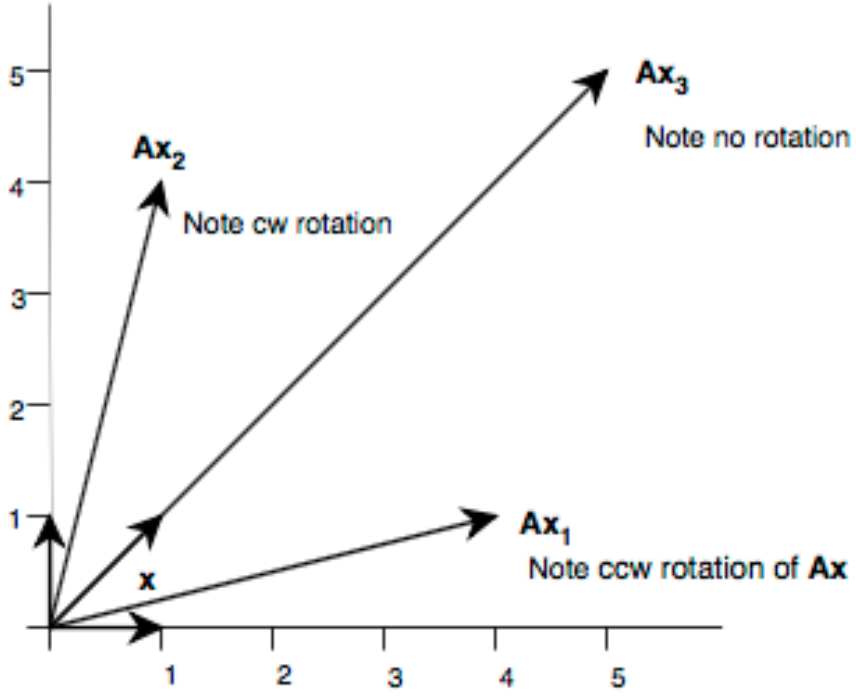
假设吸毒者每次检测呈阳性（+）的概率为99%。而不吸毒者每次检测呈阴性（-）的概率为99%。假设某公司对全体雇员进行吸毒检测，已知0.5%的雇员吸毒。请问每位检测结果呈阳性的雇员吸毒的概率有多高？（作业，答案0.3322）

概率论总结(板书)

方阵的特征值(Eigenvalues)与特征向量(Eigenvectors)

$\mathbf{Ax}=\lambda\mathbf{x}$几何意义，并思考如何计算$\mathbf{A}^{1000}$

给定一个矩阵$\mathbf{A}=\begin{bmatrix}4 & 1\\ 1 & 4\end{bmatrix}$，对于$\mathbf{x}_1=\begin{bmatrix}1\\ 0\end{bmatrix}$,则有$\mathbf{Ax}_1=\begin{bmatrix}4\\ 1\end{bmatrix}$；对于$\mathbf{x}_3=\begin{bmatrix}1\\ 1\end{bmatrix}$，则有$\mathbf{Ax}_3=5\begin{bmatrix}1\\ 1\end{bmatrix}$



特征分解的性质

特征分解的一般性质

- 对于$\mathbf{A}\mathbf{x}_i = \lambda \mathbf{x}_i$，如果所有的特征值都不相同，则相应的所有的特征向量线性无关。此时，$\mathbf{A}$可以被对角化为

$$\mathbf{A} = \mathbf{V}\Lambda\mathbf{V}^{-1}. \tag{11}$$

其中$\mathbf{V} = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$，$\Lambda = \mathrm{Diag}(\lambda_1, \ldots, \lambda_n)$。思考$\mathbf{A}^{1000}$。

- 思考：所有的方阵都可以对角化吗?

对称矩阵的特征分解(1/2)

- 如果一个对称矩阵的特征值不同，则其相应的所有的特征向量正交($\mathbf{U}\mathbf{U}^T = \mathbf{U}^T\mathbf{U} = \mathbf{I}$)

$$\mathbf{A} = \mathbf{U}\Lambda\mathbf{U}^T \tag{12}$$

$$= [\mathbf{u}_1, \ldots, \mathbf{u}_n] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{bmatrix} \begin{bmatrix} \mathbf{u}_1^T \\ \vdots \\ \mathbf{u}_n^T \end{bmatrix} \tag{13}$$

$$= \sum_{i=1}^{n} \lambda_i \mathbf{u}_i \mathbf{u}_i^T \tag{14}$$

- 深入思考公式(14)
- 思考，如果一个对称矩阵的特征值相同，是否也可以找到相互正交的特征向量?
- 思考实际工程中对称矩阵多吗?

对称矩阵的特征分解(2/2)

- 对称矩阵的特征值是实数
- 如果$\mathbf{A} \in \mathbf{R}^{n \times n}$是一对称矩阵且rank $r \leq n$，则有

$$\underbrace{|\lambda_1| \geq |\lambda_2| \geq \ldots \geq |\lambda_r|}_{r} > \underbrace{\lambda_{r+1} = \ldots \lambda_n}_{n-r} = 0 \tag{15}$$

- $\mathrm{Rank}\left(\mathbf{A}^T\mathbf{A}\right) = \mathrm{Rank}\left(\mathbf{A}\mathbf{A}^T\right) = \mathrm{Rank}\left(\mathbf{A}\right) = \mathrm{Rank}\left(\Lambda\right)$
- 思考对于任意矩阵，能否找到一个类似的分解?

二次型(Quadratic Form)

- 给定矩阵$\mathbf{A} \in \mathbb{R}^{n \times n}$，函数

$$\mathbf{x}^T \mathbf{A}\mathbf{x} = \sum\sum x_i x_j a_{ij} \tag{16}$$

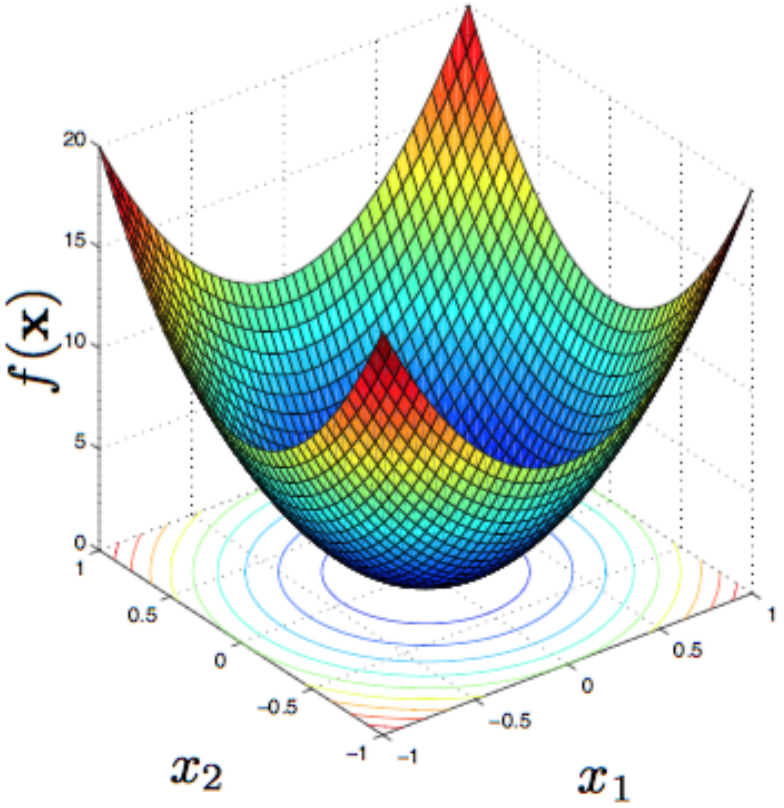
被称为二次型。

- 如果对于所有$\mathbf{x} \in \mathbb{R}^n$，有$\mathbf{x}^T \mathbf{A}\mathbf{x} \geq 0$，则为半正定矩阵(positive semidefinite)，此时$\lambda(\mathbf{A}) \geq 0$。
- 如果对于所有$\mathbf{x} \in \mathbb{R}^n, \mathbf{x} \neq \mathbf{0}$，有$\mathbf{x}^T \mathbf{A}\mathbf{x} > 0$，则为正定矩阵(positive definite)。
- 负定矩阵
- 不定矩阵(indefinite)

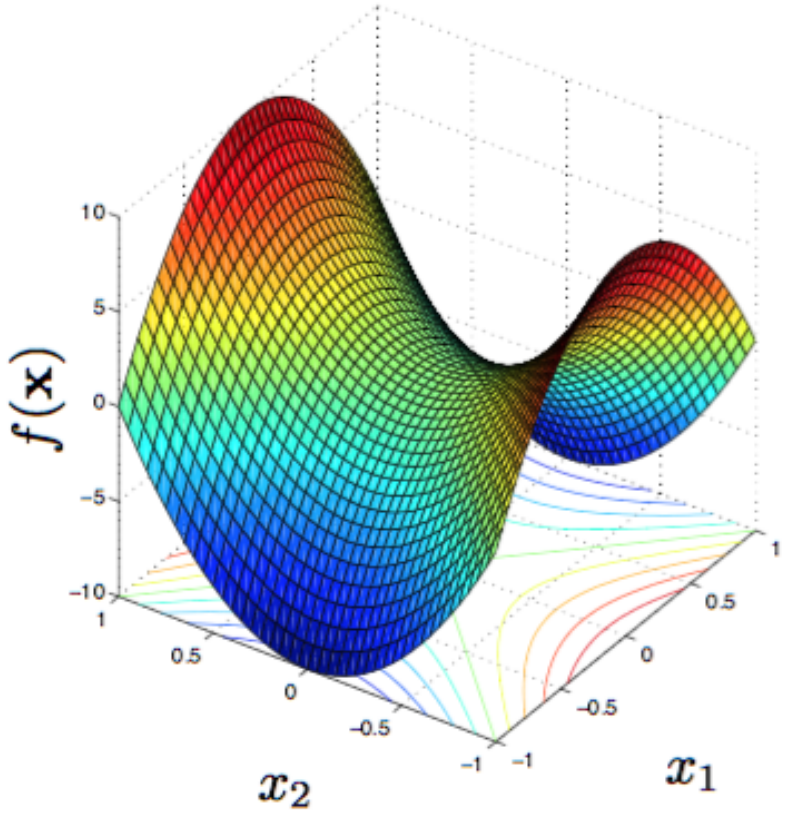
二次型

二次型图形

- 二次函数$f(\mathbf{x}) = \mathbf{x}^T\mathbf{A}\mathbf{x} + 2\mathbf{b}^T\mathbf{x} + c$
- $\nabla f(\mathbf{x}) = 2\mathbf{A}\mathbf{x} + 2\mathbf{b}$, $\nabla^2 f(\mathbf{x}) = 2\mathbf{A}$



(a) PSD A.



(b) indefinite A.

特征分解的应用—PCA本质讲述(1/3)

PCA的本质(协方差矩阵的相似对角化，KL变换)

- 给定一个矩阵$\mathbf{X} \in \mathbb{R}^{m \times n}$，例如

$$\mathbf{X} = \begin{bmatrix} a_1 & a_2 & \cdots & a_n \\ b_1 & b_2 & \cdots & b_n \end{bmatrix}$$

选择$k < m$个正交基进行降维的同时又尽量保留原始的信息。即，使得$\mathbf{A}$变换到这组基后，使得行向量间的协方差为0，而每个行向量的方差尽可能大。

- 协方差矩阵(对称半正定)为

$$\mathbf{C}_X = \frac{1}{n}\mathbf{X}\mathbf{X}^T = \begin{bmatrix} \frac{1}{n}\sum_{i=1}^n a_i^2 & \frac{1}{n}\sum_{i=1}^n a_i b_i \\ \frac{1}{n}\sum_{i=1}^n a_i b_i & \frac{1}{n}\sum_{i=1}^n b_i^2 \end{bmatrix}$$

特征分解的应用—PCA本质讲述(2/3)

PCA的本质

- 问题：假设变换矩阵为$\mathbf{Y}=\mathbf{QX}$，并先假设$\mathbf{Q}$是方阵(先不降维)，则有

$$\mathbf{C}_Y=\frac{1}{n}\mathbf{Y}\mathbf{Y}^T=\mathbf{Q}\mathbf{C}_X\mathbf{Q}^T$$

  如何使得$\mathbf{C}_Y$是一个对角矩阵？回忆$\mathbf{C}_X=\mathbf{U}\Lambda\mathbf{U}^T\Rightarrow\Lambda=\mathbf{U}^T\mathbf{C}_X\mathbf{U}$. 如果$\mathbf{Q}=\mathbf{U}^T$?

- 思考如何降维?

特征分解的应用—PCA本质讲述(3/3)

PCA降维举例(思考特征值反应了什么？)

1. $\mathbf{X}=\begin{bmatrix}-1 & -1 & 0 & 2 & 0\\ -2 & 0 & 0 & 1 & 1\end{bmatrix}$, $\mathbf{C}_X=\begin{bmatrix}\frac{6}{5} & \frac{4}{5}\\ \frac{4}{5} & \frac{6}{5}\end{bmatrix}$
2. 计算$\mathbf{C}_X$特征值为：$\lambda_1=2,\ \lambda_2=2/5$, 特征值特征向量为$\begin{bmatrix}\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix},\begin{bmatrix}-\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}}\end{bmatrix}$, 因此$\mathbf{U}=\begin{bmatrix}\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$,则$\mathbf{U}^T=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$, 此时如果$\mathbf{Q}=\mathbf{U}^T=\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}$，则只对角化了$\mathbf{C}_Y$，未降维。降维则是取$\mathbf{Q}$的第一行
3. 降维：$\begin{bmatrix}\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}}\end{bmatrix}\mathbf{X}=\begin{bmatrix}-\frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 & \frac{3}{\sqrt{2}} & -\frac{1}{\sqrt{2}}\end{bmatrix}$，此时可验证$\mathbf{C}_Y=2=\lambda_1$

矩阵总结(板书)

一般约束优化问题

- 约束优化问题一般形式：

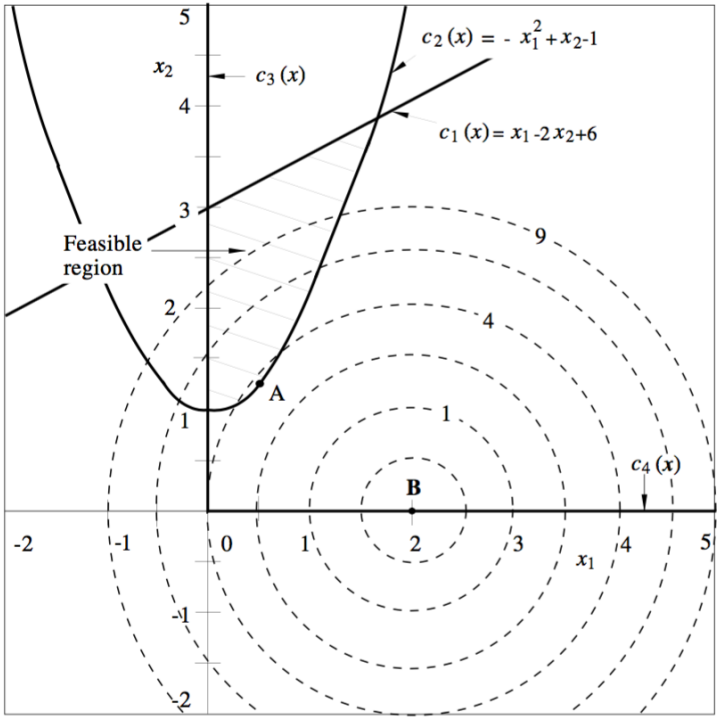
$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) \\ \text{subject to} \quad & a_i(\mathbf{x}) = 0 \text{ for } i = 1, 2, \cdots, p \\ & c_j(\mathbf{x}) \geq 0 \text{ for } j = 1, 2, \cdots, q \end{aligned} \tag{17}$$

- 可行域：满足$f(\mathbf{x})$定义域和约束条件的$\mathbf{x}$的集合。$c_j(\mathbf{x}) = 0$表明不等式约束被激活(active)。

一般约束优化问题(举例)

- 考虑以下约束优化问题

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) = x_1^2 + x_2^2 - 4x_1 + 4 = (x_1 - 2)^2 + x_2^2 \\ \text{subject to} \quad & c_1(\mathbf{x}) = x_1 - 2x_2 + 6 \geq 0 \\ & c_2(\mathbf{x}) = -x_1^2 + x_2 - 1 \geq 0, c_3(\mathbf{x}) = x_1 \geq 0, c_4(\mathbf{x}) = x_2 \geq 0 \end{aligned}$$



一般约束优化问题极值点一阶必要条件(Karush-Kuhn-Tucker(KKT))

KKT条件(思考如何把约束优化转化为无约束优化)

- 如果$\mathbf{x}^*$是约束优化问题的局部最小解，那么有
  1. $a_i(\mathbf{x}^*) = 0$ for $i = 1, 2, \cdots, p$
  2. $c_j(\mathbf{x}^*) \geq 0$ for $j = 1, 2, \cdots, q$
  3. 存在Lagrange multipliers $\lambda_i^*, i = 1, 2, \cdots, p$和$u_j^*, j = 1, 2, \cdots, p$ 使得

$$\nabla f(\mathbf{x}^*) = \sum_{i=1}^{p} \lambda_i^* \nabla a_i(\mathbf{x}^*) + \sum_{j=1}^{q} u_j^* \nabla c_j(\mathbf{x}^*)$$

  4. $u_j^* c_j(\mathbf{x}^*) = 0$ for $j = 1, 2, \cdots, q$
  5. $u_j^* \geq 0$ for $j = 1, 2, \cdots, q$
- 条件3解释：Lagrangin $L(\mathbf{x}, \boldsymbol{\lambda}, \boldsymbol{\mu}) = f(\mathbf{x}) - \sum_{i=1}^{p} \lambda_i a_i(\mathbf{x}) - \sum_{j=1}^{q} u_j c_j(\mathbf{x})$,则条件3等价于$\nabla_x L\left(\mathbf{x}^*, \boldsymbol{\lambda}^*, \boldsymbol{\mu}^*\right) = 0$

KKT应用举例(1/2)

- 函数

$$\begin{aligned} \text{minimize} \quad & f(\mathbf{x}) = x_1^2 + x_2^2 - 14x_1 - 6x_2 \\ \text{subject to} \quad & c_1(\mathbf{x}) = 2 - x_1 - x_2 \geq 0 \\ & c_2(\mathbf{x}) = 3 - x_1 - 2x_2 \geq 0 \end{aligned}$$

- KKT

$$\begin{aligned} 2x_1 - 14 + u_1 + u_2 &= 0 \\ 2x_2 - 6 + u_1 + 2u_2 &= 0 \\ u_1(2 - x_1 - x_2) &= 0 \\ u_2(3 - x_1 - 2x_2) &= 0 \\ u_1 &\geq 0 \\ u_2 &\geq 0 \end{aligned}$$

KKT应用举例(2/2)

- 解KKT条件，考虑所有的cases：不等式激活和u_i的非负性
- Case I，没有激活的情况，则$u_1^* = u_2^* = 0$，可得$x_1^* = 7, x_2^* = 3$，非解。
- Case II, 一个激活(作业)
- Case III，两个都激活(作业)

凸优化问题标准形式(Game Over)

- 凸优化问题

$$\begin{aligned} \text{minimize} \quad & f_0(\mathbf{x}) \\ \text{subject to} \quad & f_i(\mathbf{x}) \leq 0 \text{ for } i = 1, 2, \cdots, m \\ & h_i(\mathbf{x}) = 0 \text{ for } i = 1, 2, \cdots, p \end{aligned} \tag{18}$$

- 则有$f_0(\mathbf{x})$是凸函数，可行域是凸集，课上简单讲述凸函数和凸集。

优化在机器学习中的应用概述

- PCA
- ICA
- SVM，找最优的分离平面
- 线性回归
- 最大似然
- 等等

谢谢！内容多，时间有限，恳请大家批评指正！