模式识别与计算机视觉

概率、统计、线性代数极简回顾
最近邻方法的回顾

张振宇
智能科学与技术学院
2025

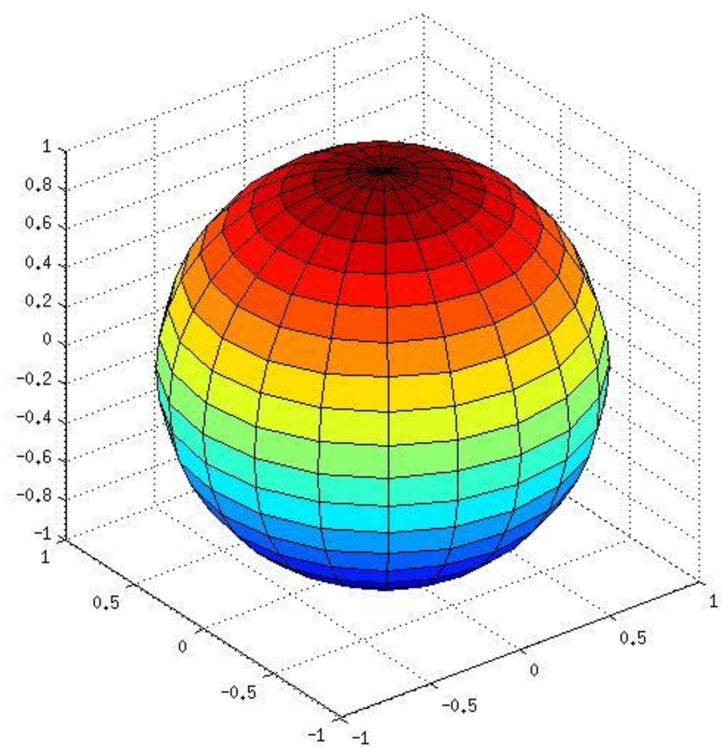
目标

- ✓ 回忆掌握相关基本概念和最重要的定理
- ✓ 能够熟练应用提供的资源查表
- ✓ 提高目标
  - 理解相关定理的证明和推导过程
  - 能不查表熟练应用重要的一些定理和推导
  - 进一步：能通过查表掌握一些课堂没有讲授的定理，并能应用到学习、研究中遇到的问题中去

线性代数

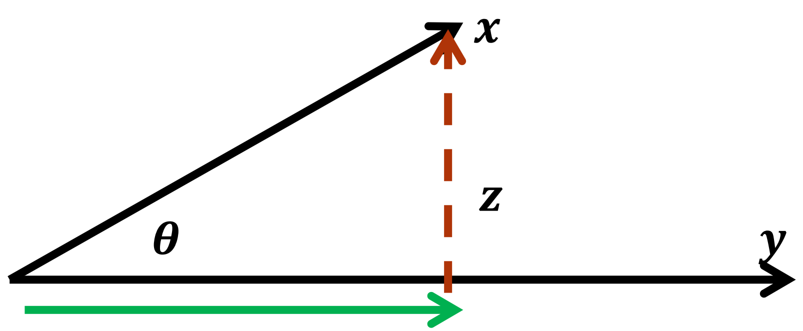
向量 (vector)

- ✓ $\boldsymbol{x} = (x_1, x_2, \ldots, x_d)^T \in \mathbb{R}^d$
- ✓ 内积 (dot-product, inner-product,点积)
  - $\boldsymbol{x}^T\boldsymbol{y} = \boldsymbol{y}^T\boldsymbol{x} = \sum_{i=1}^{d} x_i y_i$
- ✓ 向量的长度(vector norm)
  - $\|\boldsymbol{x}\| = \sqrt{\boldsymbol{x}^T\boldsymbol{x}}, \|\boldsymbol{x}\|^2 = \boldsymbol{x}^T\boldsymbol{x}$
  - 若$\|\boldsymbol{x}\| = 1$，称$\boldsymbol{x}$为单位向量
- ✓ 正交(orthogonal)
  - $\boldsymbol{x}^T\boldsymbol{y} = 0$
  - $\boldsymbol{x}$和$\boldsymbol{y}$被称为垂直 (perpendicular): $\boldsymbol{x} \perp \boldsymbol{y}$



内积、角度、投影

- ✓ $\boldsymbol{x}$: $\|\boldsymbol{x}\|$决定长度，$\frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}$决定方向
- ✓ 向量之间的夹角(angle):
  - $\boldsymbol{x}^T\boldsymbol{y} = \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\| \cdot \cos\theta$
  - $\boldsymbol{x}^T\boldsymbol{y} \le \|\boldsymbol{x}\| \cdot \|\boldsymbol{y}\|$
- ✓ $\boldsymbol{x}$在$\boldsymbol{y}$上的投影(projection)
  - 方向：$\frac{\boldsymbol{y}}{\|\boldsymbol{y}\|}$　　长度：$\|\boldsymbol{x}\|\cos\theta = \|\boldsymbol{x}\|\frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{x}\|\cdot\|\boldsymbol{y}\|} = \frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{y}\|}$
  - 投影$\mathrm{proj}_{\boldsymbol{y}}\boldsymbol{x}$：$\frac{\boldsymbol{x}^T\boldsymbol{y}}{\|\boldsymbol{y}\|^2}\boldsymbol{y}$
  - $\mathrm{proj}_{\boldsymbol{y}}\boldsymbol{x} \perp \boldsymbol{z}$
  - $\mathrm{proj}_{\boldsymbol{y}}\boldsymbol{x} + \boldsymbol{z} = \boldsymbol{x}$



柯西-施瓦茨不等式

✓ Cauchy's inequality

- $\left(\sum_{k=1}^{n} a_k b_k\right)^2 \leq \left(\sum_{k=1}^{n} a_k^2\right)\left(\sum_{k=1}^{n} b_k^2\right)$
- 等号成立当且仅当存在固定实数c，使得$\forall k$，$a_k = cb_k$

✓ Schwarz's Inequality

- $\left[\int_a^b f(x)g(x)\mathrm{d}x\right]^2 \leq \left[\int_a^b [f(x)]^2\mathrm{d}x\right]\left[\int_a^b [g(x)]^2\mathrm{d}x\right]$
- 等号成立当且仅当存在固定实数c，使得$\forall x \in [a\ b]$，$f(x) = cg(x)$

矩阵(Matrix)

- ✓ $X = \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ \vdots & \ddots & \vdots \\ x_{m1} & \cdots & x_{mn} \end{bmatrix}$: $m \times n$的矩阵
  - $n = m$时称为方阵(square matrix)
  - 行矩阵(row matrix，行向量): $m = 1$
  - 列矩阵(column matrix，列向量，向量): $n = 1$

- ✓ 对角阵(diagonal matrix)：方阵中，只有对角线非零
- ✓ 单位阵(identity matrix)：对角线全部为1的对角阵
  - 一般记为I或者I_n

矩阵运算

- ✓ 乘法：$X: m \times n, Y: n \times p$
  - 维度(dimensionality)相符时乘法才有定义
  - 一般来说$XY \neq YX$
- ✓ 矩阵的幂(power)
  - 对方阵有定义：$X^2 = XX, X^3 = XXX, \ldots$
- ✓ 转置(transpose)
  - $X: m \times n$，那么 $X^T: n \times m$
  - $X^T X: n \times n, XX^T: m \times m$
- ✓ 对称矩阵(symmetric matrix)
  - 是方阵, $X_{ij} = X_{ji}, \ \forall i, j$

行列式值、矩阵的逆

✓ 方阵的行列式值(determinant)

- $|X|$, 或写作$\det(X)$
- $|X|=|X^T|$
- $|XY| = |X||Y|$
- $|\lambda X| = \lambda^n|X| \quad (X: n \times n)$

✓ 方阵的逆矩阵(inverse matrix)

- X^{-1}: 满足$XX^{-1} = X^{-1}X = I_n$
- X可逆(invertible) $\equiv$ $|X| \neq 0 \quad (\Leftrightarrow)$
- $(X^{-1})^{-1} = X, (\lambda X)^{-1} = \frac{1}{\lambda}X^{-1}$
- $(XY)^{-1} = Y^{-1}X^{-1}, (X^{-1})^T = \left(X^T\right)^{-1}$

方阵的特征值、特征向量、迹

- 特征值(eigenvalue)和特征向量(eigenvector)
  - $A\boldsymbol{x} = \lambda\boldsymbol{x} \qquad A: n \times n$
  - λ:特征值 $\boldsymbol{x}$:特征向量
- n阶方阵有n个特征值
  - 可能存在相等的特征值
- 特征值和对角线的关系
  - $\sum_{i=1}^{n} a_{ii} = \sum_{i=1}^{n} \lambda_i$
  - $\det(A) = \prod_{i=1}^{n} \lambda_i$
- 方阵的迹(trace)
  - $\mathrm{tr}(A) = \sum_{i=1}^{n} a_{ii} = ??$, $\mathrm{tr}(AB) = \mathrm{tr}(BA)$

实对称矩阵

- ✓ 对称矩阵，每个项都是实数
  - Real symmetric matrix
  - 这门课程中最常用到
- ✓ 性质：
  - 所有特征值都是实数，特征向量都是实向量
  - 特征值记为$\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$
  - 对应的特征向量记为$\xi_1, \xi_2, \ldots, \xi_n$，单位向量
  - 特征向量互相垂直：$\xi_i^T \xi_j = 0 \ \ (i \neq j)$
  - $E = [\xi_1 \ \xi_2 \ \ \ldots \ \xi_n]$是$n \times n$的，是满秩(full rank)的，$\operatorname{rank}(E) = n$.

实对称矩阵的分解(decomposition)

- ✓ $X: n \times n$的实对称矩阵
  - 特征值为λ_i，其对应的特征向量为ξ_i
- ✓ $X = \sum_{i=1}^{n} \lambda_i \xi_i \xi_i^T$
  - 称为谱分解(spectral decomposition)
  - 约定$\|\xi_i\| = 1$, 则E是正交矩阵(orthogonal matrix)
  - $X = E \Lambda E^T$
    - Λ是一个对角矩阵，$\Lambda_{ii} = \lambda_i$
    - $EE^T = E^T E = I$, $E^{-1} = ?$, $|E| = ?$
- ✓ 进一步阅读
  - LU分解, Cholesky分解, QR分解
  - 资源： Numerical Recipes series

正定、半正定

- 对称方阵A是正定的(positive definite)当且仅当
  - $\forall \boldsymbol{x} \neq \boldsymbol{0} \quad \boldsymbol{x}^T A \boldsymbol{x} = \sum_{i,j} x_i x_j A_{ij} > 0$
  - $\forall \boldsymbol{x} \neq \boldsymbol{0} \quad \boldsymbol{x}^T A \boldsymbol{x} \geq 0$ 则A为半正定(positive semi-definite)
  - 分别记为$A \succ 0$或$A \succcurlyeq 0$
- $\boldsymbol{x}^T A \boldsymbol{x}$: 称为二次型(quadratic)
  - 这门课程会经常用到，一般满足$A \succcurlyeq 0$
- 等价关系
  - 1. $A \succ 0$ ($A \succcurlyeq 0$)
  - 2. 特征值全部为正数 (非负实数)
- 正定矩阵的任意主子矩阵也是正定矩阵

矩阵求导

- 假设一切求导的条件都满足（导数都存在）
- $\frac{\partial \boldsymbol{a}}{\partial x}$是一个向量，$\left(\frac{\partial \boldsymbol{a}}{\partial x}\right)_i = \frac{\partial a_i}{\partial x}$
- 对于矩阵， $\left(\frac{\partial A}{\partial x}\right)_{ij} = \frac{\partial A_{ij}}{\partial x}$
- $\left(\frac{\partial x}{\partial \boldsymbol{a}}\right)_i = \frac{\partial x}{\partial a_i}$ $\quad\left(\frac{\partial x}{\partial A}\right)_{ij} = \frac{\partial x}{\partial A_{ij}}$ $\quad\left(\frac{\partial \boldsymbol{a}}{\partial \boldsymbol{x}}\right)_{ij} = \frac{\partial a_i}{\partial x_j}$
- 如何求导？
  - 能够查表并合理应用
  - The Matrix Cookbook 最好打印前两章上课时带着

矩阵求导

2 Derivatives

This section is covering differentiation of a number of expressions with respect to a matrix $\mathbf{X}$. Note that it is always assumed that $\mathbf{X}$ has *no special structure*, i.e. that the elements of $\mathbf{X}$ are independent (e.g. not symmetric, Toeplitz, positive definite). See section 2.8 for differentiation of structured matrices. The basic assumptions can be written in a formula as

$$\frac{\partial X_{kl}}{\partial X_{ij}} = \delta_{ik}\delta_{lj} \tag{32}$$

that is for e.g. vector forms,

$$\left[\frac{\partial \mathbf{x}}{\partial y}\right]_i = \frac{\partial x_i}{\partial y} \qquad \left[\frac{\partial x}{\partial \mathbf{y}}\right]_i = \frac{\partial x}{\partial y_i} \qquad \left[\frac{\partial \mathbf{x}}{\partial \mathbf{y}}\right]_{ij} = \frac{\partial x_i}{\partial y_j}$$

The following rules are general and very useful when deriving the differential of an expression ([19]):

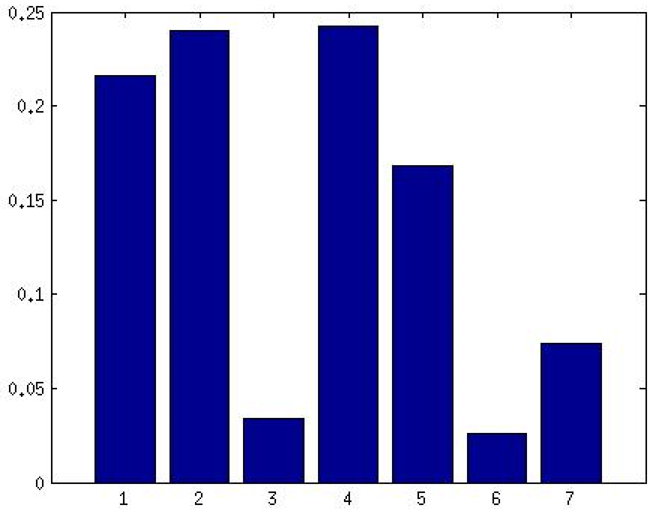
$$\begin{aligned}
\partial \mathbf{A} &= 0 \qquad (\mathbf{A} \text{ is a constant}) && (33)\\
\partial(\alpha\mathbf{X}) &= \alpha\partial\mathbf{X} && (34)\\
\partial(\mathbf{X}+\mathbf{Y}) &= \partial\mathbf{X}+\partial\mathbf{Y} && (35)\\
\partial(\mathrm{Tr}(\mathbf{X})) &= \mathrm{Tr}(\partial\mathbf{X}) && (36)\\
\partial(\mathbf{XY}) &= (\partial\mathbf{X})\mathbf{Y}+\mathbf{X}(\partial\mathbf{Y}) && (37)\\
\partial(\mathbf{X}\circ\mathbf{Y}) &= (\partial\mathbf{X})\circ\mathbf{Y}+\mathbf{X}\circ(\partial\mathbf{Y}) && (38)\\
\partial(\mathbf{X}\otimes\mathbf{Y}) &= (\partial\mathbf{X})\otimes\mathbf{Y}+\mathbf{X}\otimes(\partial\mathbf{Y}) && (39)\\
\partial(\mathbf{X}^{-1}) &= -\mathbf{X}^{-1}(\partial\mathbf{X})\mathbf{X}^{-1} && (40)\\
\partial(\det(\mathbf{X})) &= \mathrm{Tr}(\mathrm{adj}(\mathbf{X})\partial\mathbf{X}) && (41)\\
\partial(\det(\mathbf{X})) &= \det(\mathbf{X})\mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (42)\\
\partial(\ln(\det(\mathbf{X}))) &= \mathrm{Tr}(\mathbf{X}^{-1}\partial\mathbf{X}) && (43)\\
\partial\mathbf{X}^T &= (\partial\mathbf{X})^T && (44)\\
\partial\mathbf{X}^H &= (\partial\mathbf{X})^H && (45)
\end{aligned}$$

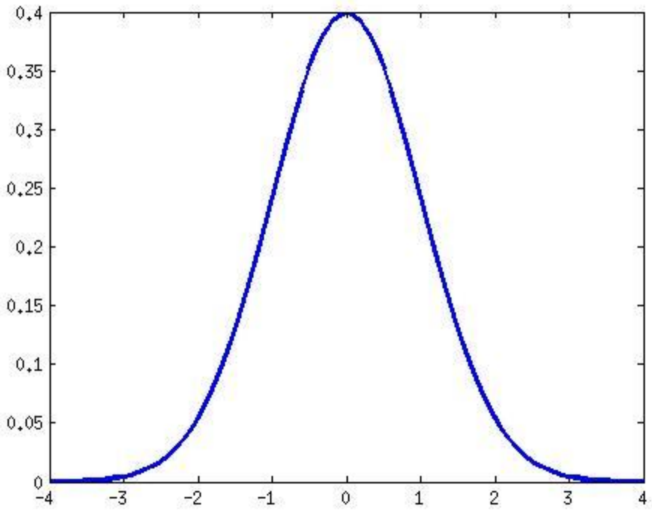
概率与统计

Probability & Statistics

随机变量(Random variable)

- ✓ X: 可以是离散(discrete)、连续(continuous)、或者混合(hybrid)的





概率密度函数

- （古典）离散(discrete)：
  - 可数的(countable)不相容的若干事件$c_1, c_2, \dots$
  - $p(X = x_i) = c_i$ -- probability mass function
  - $c_i \geq 0$， $\sum_i c_i = 1$ --pmf是合法的
- 连续(continuous)：为简化，只考虑$X \in (-\infty, \infty)$
  - $p(x)$:概率密度函数probability density function (pdf)
  - $p(x) \geq 0$, $\int_{-\infty}^{+\infty} p(x)\,\mathrm{d}x = 1$
- 随机变量可以粗略地看成是一个函数，而不是一个数学分析意义上的变量

累积分布函数(连续)

- ✓ Cumulative distribution function （cdf）
- ✓ $F(x) = P(X \leq x) = \int_{-\infty}^{x} p(x)\,\mathrm{d}x$
  - $F(-\infty) = 0 \leq F(x) \leq F(+\infty) = 1$
  - 非减性(non-decreasing)

    如果$x \leq y$，那么$F(x) \leq F(y)$
  - $P(X = x) =$?
  - $P(x_1 < x < x_2) = F(x_2) - F(x_1) = \int_{x_1}^{x_2} p(x)\,\mathrm{d}x$
- ✓ PDF和CDF的关系
  - $p(x) = F'(x)$

联合、条件分布、变换

- 联合(joint distribution)： $P(X=\boldsymbol{x})$ 多维变量
  - $p(\boldsymbol{x}) \geq 0 \qquad \int p(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 1$
- 条件(conditional distribution): $P(X=x|Y=y)$
- $p(x,y) = p(y)p(x|y)$
- $p(x) = \int_y p(x,y)\mathrm{d}y$ --marginal（边际）分布
- 假设$x = g(y)$，那么

$$p_Y(y) = p_X(x)\left|\frac{\mathrm{d}x}{\mathrm{d}y}\right| = p_X(g(y))|g'(y)|$$

  - 对g的具体要求：单调固定函数

多维分布的期望

- ✓ 假设有函数$f(\boldsymbol{x})$，在$\boldsymbol{x}$服从分布$p(\boldsymbol{x})$时：
- ✓ f的期望(Expectation)，记为$\mathrm{E}[f(X)]$
  - $\mathrm{E}[f(X)] = \sum_{x} f(\boldsymbol{x}) \cdot p(X = \boldsymbol{x})$，或
  - $\mathrm{E}[f(X)] = \int f(\boldsymbol{x}) p(\boldsymbol{x}) \mathrm{d}\boldsymbol{x}$
  - 条件期望$\mathrm{E}(f(\boldsymbol{x})|Y = \boldsymbol{y}) = \sum_{x} f(\boldsymbol{x}) \cdot p(\boldsymbol{x}|\boldsymbol{y})$
- ✓ f的方差(Variance,一维)或协方差(covariance，多维)
  - $\mathrm{Var}(X) = \mathrm{E}[(X - EX)^2]$
  - $\mathrm{Var}(X) = \mathrm{E}[X^2] - (EX)^2$
- ✓ 当$p(\boldsymbol{x})$和$f(\boldsymbol{x})$固定时
  - 期望、方差是一个确定的数（或向量、矩阵）
  - $g(y) = E(X|Y = y)$是什么？

估计均值和协方差矩阵

- 训练样本：$\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n$
- 均值的估计estimation：

$$\overline{\boldsymbol{x}} = \frac{1}{n}\sum_{i=1}^{n} \boldsymbol{x}_i$$

- Covariance的估计

$$\mathrm{Cov}(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T$$

无偏估计unbiased estimation

$$\mathrm{Cov}(\boldsymbol{x}) = \frac{1}{n-1}\sum_{i=1}^{n}(\boldsymbol{x}_i - \overline{\boldsymbol{x}})(\boldsymbol{x}_i - \overline{\boldsymbol{x}})^T$$

两个随机变量的独立、相关

- ✓ 一般说来$p(x,y) \neq p(x)p(y)$
- ✓ 如果$\forall x, y$，$p(x,y) = p(x)p(y)$, 则X和Y互相独立(independent)
- ✓ $\mathrm{Cov}(X,Y) = \mathrm{E}[(X - EX)(Y - EY)]$
- ✓ Pearson相关系数(Pearson's correlation coefficient)：$-1 \leq \rho_{XY} \leq 1$

$$\rho_{XY} = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}},$$

  - $\rho_{XY} = 0$，称为不相关(not correlated)
  - $\rho_{XY} = \pm 1$，称为完全相关，如存在线性关系
  - 独立保证一定不相关，但是,不相关不一定能保证独立

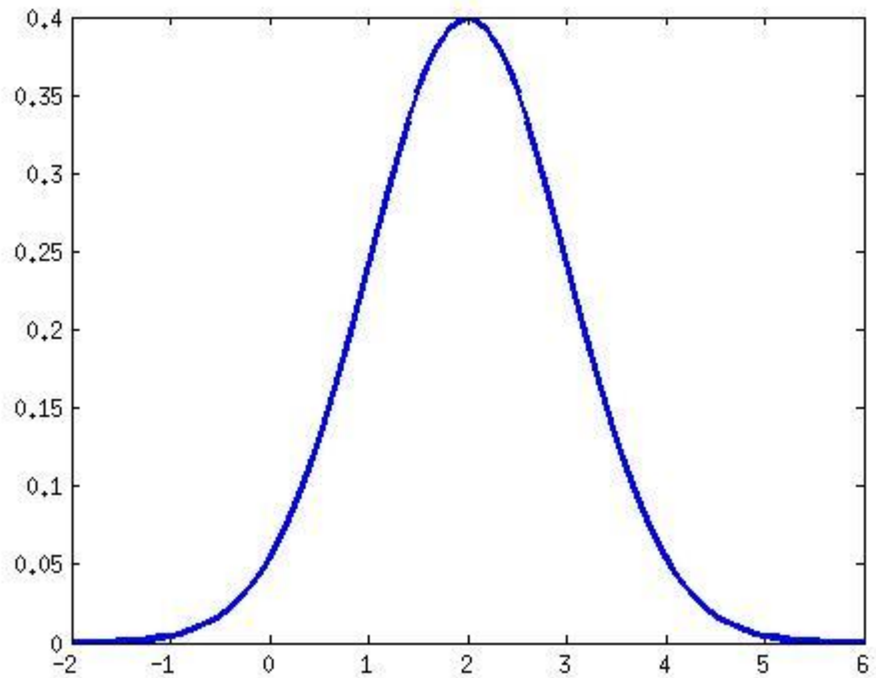
高斯分布

- 又叫正态分布,normal distribution, Gaussian distribution
- 单变量或一维高斯分布 $N\,(\mu,\sigma^2)$

$$p(x) = (2\pi)^{-\frac{1}{2}}(\sigma^2)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(x-\mu)(\sigma^2)^{-1}(x-\mu)\right\}$$

或者$\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$更眼熟？

- μ: 期望，或称均值
- σ^2:方差
  - σ:标准差(standard deviation)



- 图例中$\mu = 2, \sigma = 1$，
- Markov不等式：若$X \geq 0$(非负随机变量)，则 $P(X \geq a) \leq \frac{EX}{a}$
- Chebyshev不等式：对任何分布，$P\left((X-\mu)^2 \geq k^2\right) \leq \frac{\sigma^2}{k^2}$或$P(|X-\mu| > k) \leq \frac{\sigma^2}{k^2}$ $(k > 0)$

多维高斯分布

- 一维：

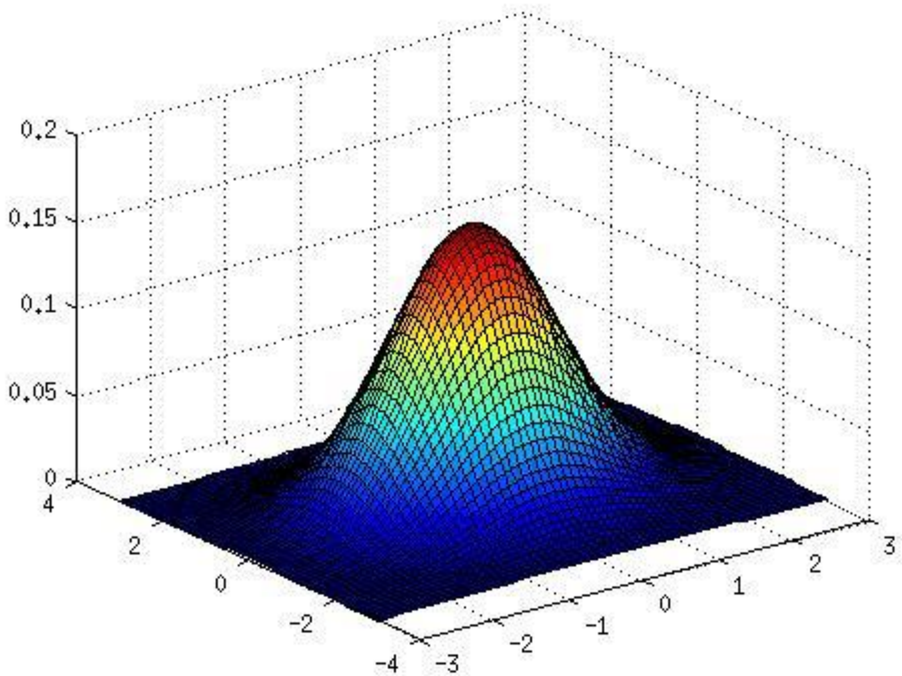
$$p(x)=(2\pi)^{-\frac{1}{2}}\left(\sigma^{2}\right)^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(x-\mu)\left(\sigma^{2}\right)^{-1}(x-\mu)\right\}$$

- 多维

$$p(\boldsymbol{x})=(2\pi)^{-\frac{D}{2}}|\Sigma|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{T}\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right\}$$

- D: 维数　　记为 $\boldsymbol{x}\sim N(\boldsymbol{\mu},\Sigma)$
- Σ: 协方差矩阵
- $\boldsymbol{\mu}$: 均值

多维高斯PDF示意图



- ✓ 图例中$\boldsymbol{\mu} = (0,0), \Sigma = I_2$
- ✓ 更多相关知识，将在PCA中讲授

高斯分布中的相关性和独立

- 一般来说，两变量
  - 独立保证一定不相关
  - 不相关不一定保证独立
- 但是，对于多维高斯分布
  - 不相关意味着协方差矩阵中非对角线项是0
  - $\begin{matrix} c_{ii} & \cdots & 0 \\ \vdots & \vdots & \vdots \\ 0 & \cdots & c_{jj} \end{matrix}$
  - 在正态分布中，不相关就等价于独立

多维与一维高斯的关系

- ✓ 多维高斯$X=\begin{pmatrix}X_a\\X_b\end{pmatrix}$
  - 条件分布：$\boldsymbol{x}_a|\boldsymbol{x}_b$ 还是高斯分布
  - 边际分布(margin distribution)：$p(\boldsymbol{x}_a)=\int p(\boldsymbol{x}_a,\boldsymbol{x}_b)\mathrm{d}\boldsymbol{x}_b$ 也是高斯分布
- ✓ 两个高斯分布的加权和也是高斯分布
  - $aX+bY$
- ✓ 为什么大家用高斯分布？

优化基础知识

- 本课程主要利用已有优化软件，不讲授优化算法或者理论
  - 最优化自身是一门复杂的课程
  - 自学教材第2.3节（到2.3.1节之前）

- 知识要点：
  - 凸集、凸函数
    - 自学教材第2.3.2节
  - 拉格朗日乘子法（等式约束）
    - 自学教材第2.3.3节
  - 算法复杂性
    - 自学教材第2.4节

进一步的阅读

- 如果对本章的内容感兴趣，可以参考如下文献
  - 教材第13章
    - 在“Advanced topics”中
    - 特点是尽量从最基本的概念出发，提供了所有必须的背景数学知识，所以会比一般的英文tutorial容易懂
- PRML的相关章节（第二章和附录）

最近邻规则

Nearest neighbor rule

问题设置problem setup

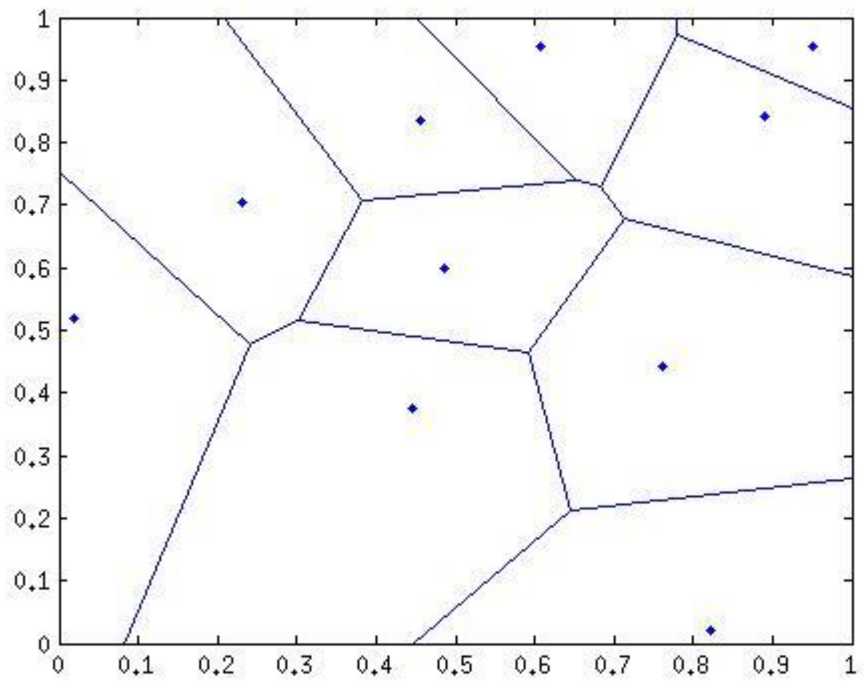
- ✓ 分类问题classification
  - 训练集$D=\{(\boldsymbol{x}_1,y_1),(\boldsymbol{x}_2,y_2),\dots,(\boldsymbol{x}_n,y_n)\}$
  - 训练样本(sample)： $\boldsymbol{x}_i \in \mathcal{X} \subseteq \mathbb{R}^d$
  - 样本的标记(label)： $y_i \in \mathcal{Y}=\{1,2,\dots,C\}$
    - 样本一共被分为C个类别(category)
    - 例如，在我们的例子里，$C=2, y_i=1$（男）或者$y_i=2$（女）

- ✓ 存在一个距离(distance)函数： $d(\boldsymbol{x},\boldsymbol{y}) \in \mathbb{R}$
  - 能够度量x和y之间的距离，或者不相似程度(level of dissimilarity)

最近邻规则和Voronoi图

给定一个测试样例$\boldsymbol{x}$

1. 发现其最近邻$i^* = \underset{i}{\operatorname{argmin}}\, d(\boldsymbol{x}, \boldsymbol{x}_i)$
2. 输出对$\boldsymbol{x}$的分类的预测：y_{i^*}



Voronoi图
(Voronoi Diagram)

最近邻可能出现的问题

- ✓ 如果出现平局(tie)？
  - $d(\boldsymbol{x}, \boldsymbol{x}_i) = d(\boldsymbol{x}, \boldsymbol{x}_j)$
  - $y_i = y_j$? $y_i \neq y_j$?
- ✓ 如果出现离群点(outlier)？
  - K-近邻(kNN，k-nearest neighbor)规则
  - 可能遇到的问题？
- ✓ 能做的多好？
  - 当训练样本趋于无穷时($n \to \infty$)，最近邻的错误率最多是最佳错误率的两倍
  - 有限样本(finite sample)时的结论尚不清楚

计算、存储代价(cost)

- 假设$d(\boldsymbol{x}, \boldsymbol{y})$是欧式距离(Euclidean distance, ℓ_2 distance)
  - 其复杂度(complexity)是$O(d)$
  - NN的复杂度$O(nd)$
    - DHS 152页的复杂度是错的
  - K-NN的复杂度同样是$O(nd)$
    - 或者是$O(nd) + O(n) + O(k)$，但通常k较小，可以忽略
    - 从n个数（距离）中选择k个最小的，复杂度是？

- 考虑一下，如果是ILSVRC，需要多长时间，多大的存储空间？这是NN的主要问题
    - $n = 1{,}200{,}000$
    - $d = 262{,}144$

降低NN的计算、存储代价

- ✓ 近似最近邻(approximate nearest neighbor， ANN)
  - 不要求一定是距离最短的k个
  - 如第k个NN，其距离是d_k，则ANN要求其选取的所有k个样例的距离$\hat{d}$满足$\hat{d} \le (1+\epsilon)d_k$即可
  - 可以将kNN搜索(search、查找)速度提高几个数量级

- ✓ 二值哈希(binary hashing)
  - hash函数f_i:将$\mathbb{R}^d$分为两部分，分别用$f_i = 0$，1表示
  - 设计m个hash函数$f_1, \dots, f_m$，每个$\boldsymbol{x}$表示为m个bit
  - $m \ll d$，计算和存储大幅简化，需要设计好的hash