复习与考试

§1 复习什么
§2 怎么复习
§3 复习举例

考查什么

- 基本概念的理解
  - 填空题、判断题、选择题
  - 例如：经典的集成学习方法等
- 重要公式的计算
  - 选择题、简答题、计算题
  - 例如：反向传播梯度计算等
- 经典算法的应用
  - 计算题
  - 例如：LDA计算、最大似然估计、贝叶斯估计等

复习什么

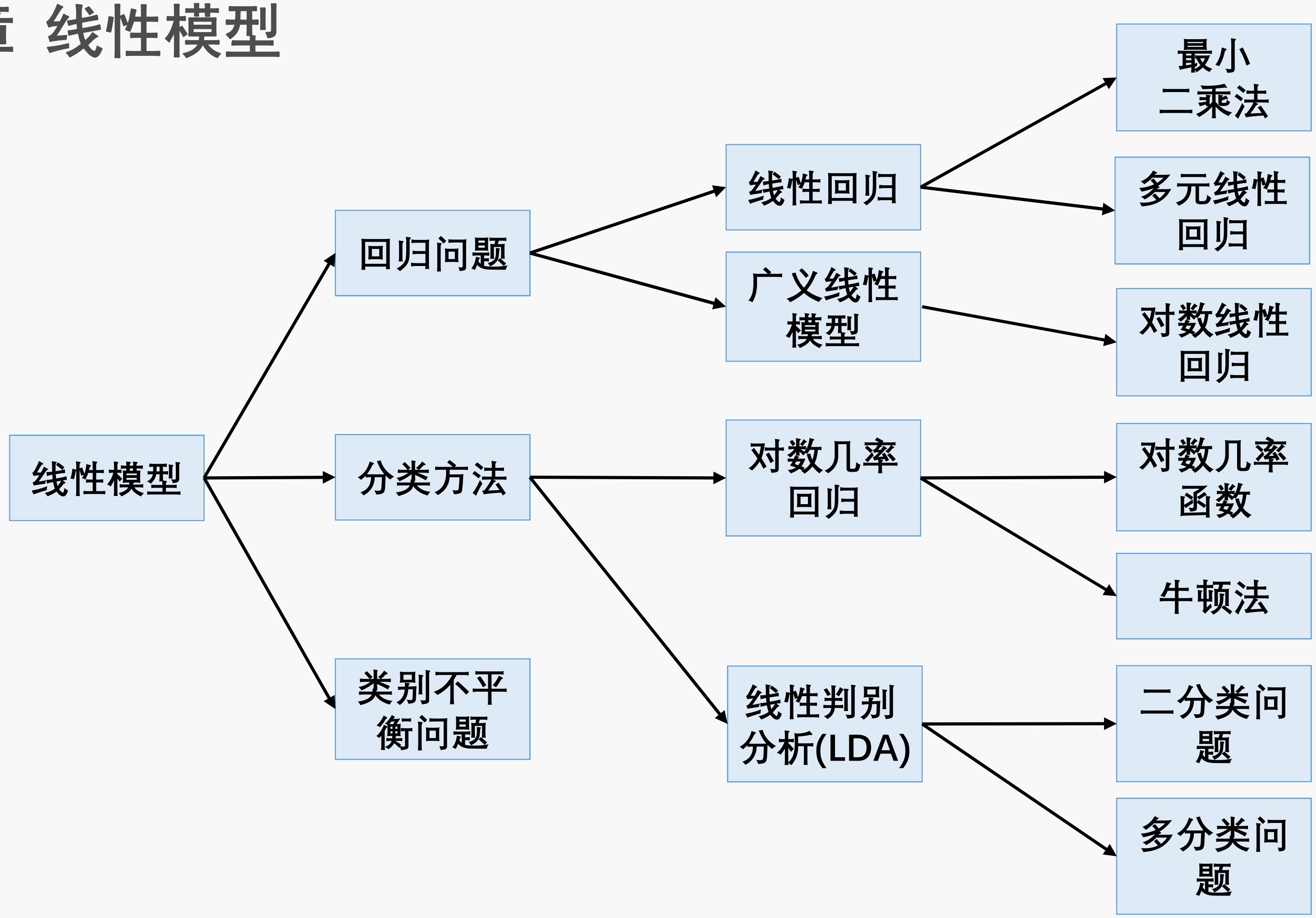
- **课程课件**
  - **例题**必须熟练掌握
  - **常见题型**的知识点需要理解透彻
- **课程作业**
  - **概念题**，**计算证明题**必须熟练掌握
  - 编程题不做要求
- **知识范畴**
  - 考试的**计算题**、**证明题**所需知识点会被课件例题、作业题基本覆盖
  - 考试的**概念题**范围较广，常见题型部分的知识点均可能进行考察

怎样复习

- 注意体会每种方法的**核心思想**，以及方法间的**逻辑关系**
  - 以特征选择为例
    - 过滤式特征选择方法主张“先选特征，再训练学习器”；而包裹式特征选择方法主张“每选择一组特征都训练一次学习器”。
    - 过滤式特征选择胜在快速，但根据统计指标选择的特征不一定适合学习器；包裹式特征选择直接使用学习器性能衡量每组特征，但开销更大
- 对于具体算法，在**理解算法步骤**的基础上，体会其**物理意义**
  - 以支持向量机为例
    - 掌握原问题、对偶问题的数学推导
    - 通过画图理解支持向量的重要性

复习举例

- 第二章 线性模型



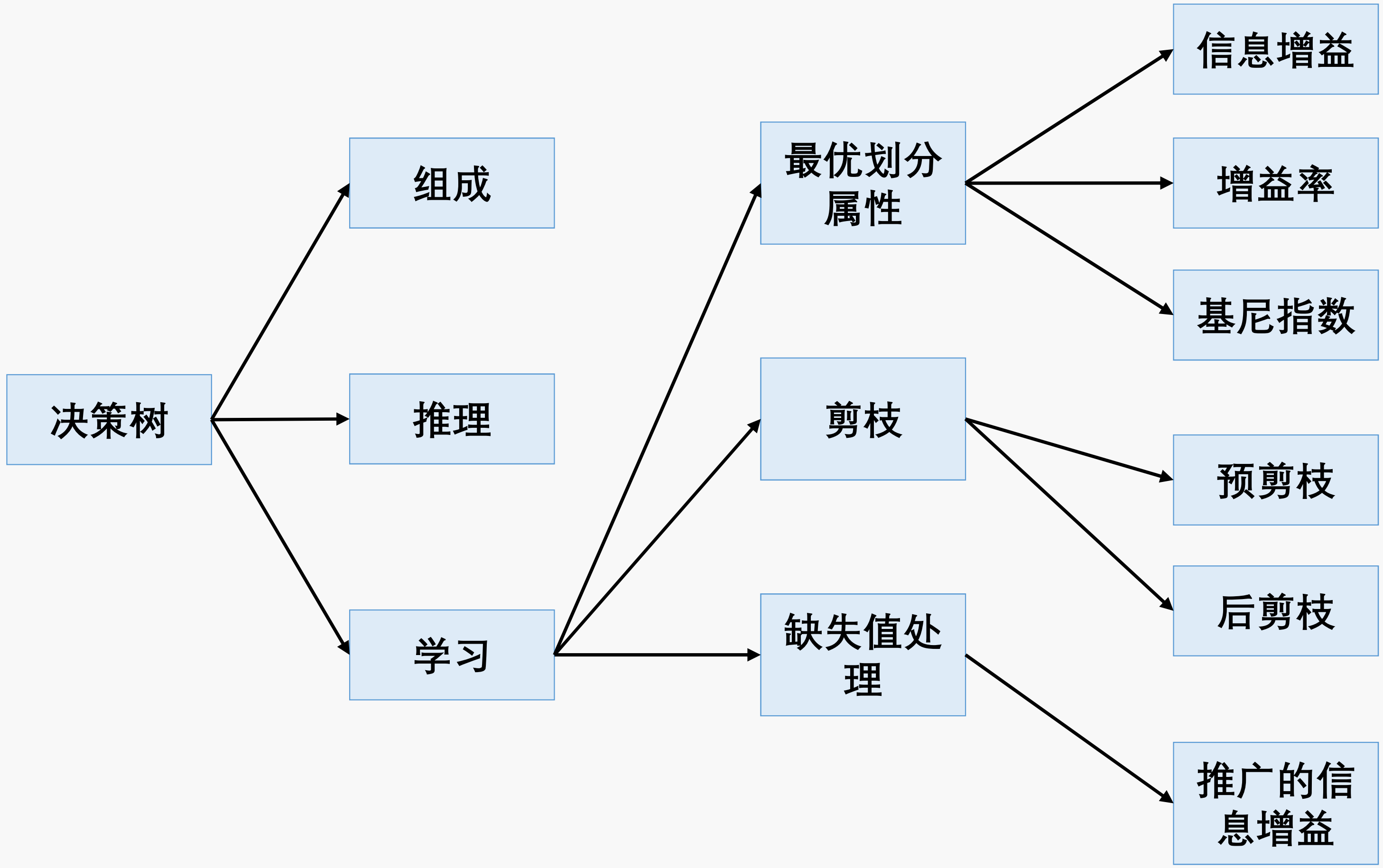
重要知识点

第二章 线性模型

- 线性回归计算
  - 应用最小二乘法进行线性回归参数估计
  - 最小二乘法推广到多元线性回归的矩阵形式
- 广义线性模型
  - 基本概念与定义
  - 对数线性回归与广义线性模型的关系
- 对数几率回归
  - 利用极大似然法的优化目标推导、求解过程
- 线性判别分析计算
  - 类内散度矩阵、类间散度矩阵、LDA优化目标推导与求解

复习举例

第三章 决策树



重要知识点

- 第三章 决策树
  - 推理
    - 根据训练集确定叶子节点的标签，计算验证集准确率
  - 最优划分属性
    - 三种准则的表达式及其物理含义
    - 根据任意一种准则构建决策树
  - 剪枝
    - 两种剪枝策略的操作过程及其特点
    - 运用任意剪枝策略构建决策树
  - 缺失值处理
    - 根据推广的信息增益准则处理简单的缺失值问题

典型例题

- 下表由15个样本组成的贷款申请数据集，包括申请人的年龄、收入情况、是否有车、信用情况等四项属性，最后一列为是否同意贷款作为预测结果

问题：假如我们按照小于30岁、30到60岁、大于60岁将申请人的年龄分为三组。以红框中的样本为验证集，其它数据为训练集，使用信息增益准则和后剪枝策略生成决策树。

序号	年龄	是否有车	收入情况	信用情况	是否同意贷款
1	19	否	一般	一般	否
2	22	否	一般	好	否
3	75	否	一般	一般	否
4	21	否	一般	一般	否
5	36	否	一般	一般	否
6	40	否	一般	好	否
7	69	是	一般	好	是
8	45	是	良好	好	是
9	52	是	一般	非常好	是
10	66	是	一般	非常好	是
11	25	否	良好	好	是
12	42	是	一般	非常好	是
13	60	否	良好	好	是
14	61	否	良好	非常好	是
15	29	是	良好	一般	是

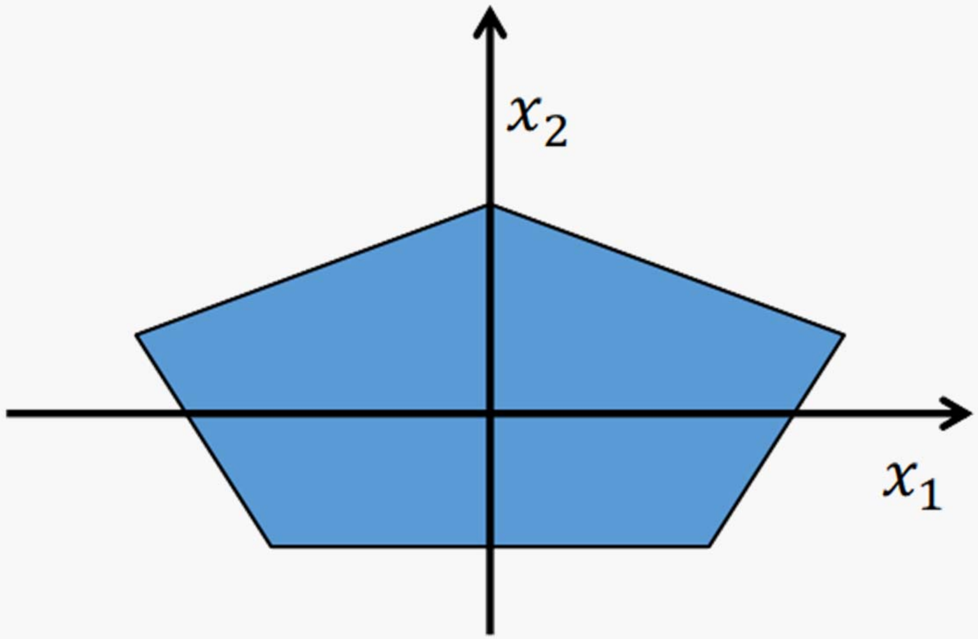
重要知识点

- 第四章 神经网络基础
  - 多层感知机
    - 了解神经元的计算方式与各种性质
    - 按要求建立能进行散点或区域二分类的多层感知机
  - BP算法
    - 能够对各种激活函数的神经元进行梯度反向传播的推导
    - 能够基于链式法则进行多层感知机参数更新的推导
  - 常见神经网络
    - 对各种常见神经网络的性质有基本的了解

典型例题

思考题

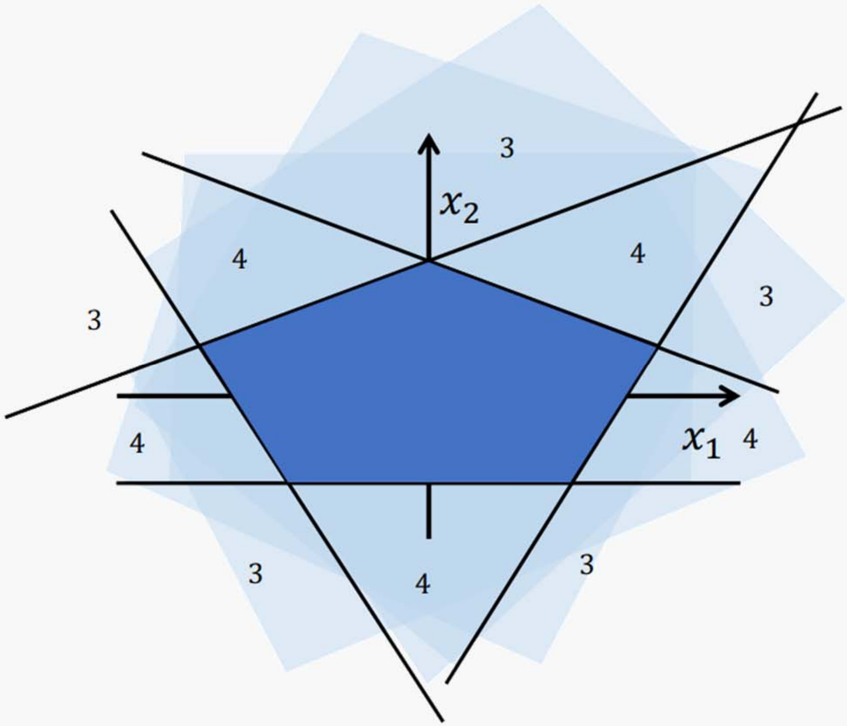
- 如何建立一个输出层神经元个数为1的多层感知机（MLP）实现下面区域的二分类建模？

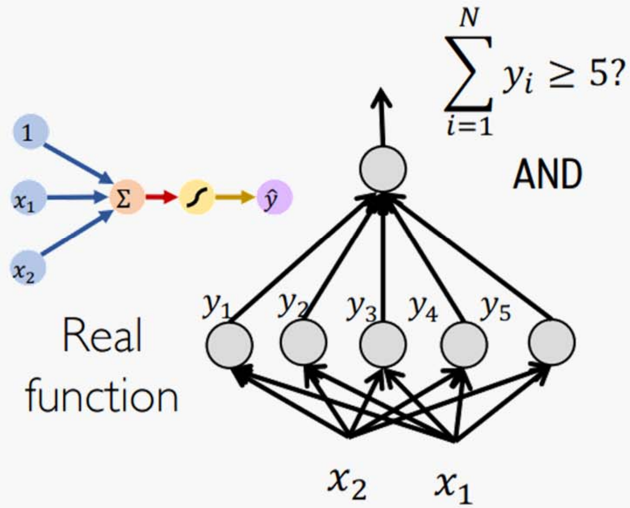


典型例题

思考题解答

- 如何建立一个输出层神经元个数为1的多层感知机（MLP）实现下面区域的二分类建模？





典型例题

反向传播算法推导

- 给定训练样本$\{x_i, y_i\}_{i=1}^n, x_i = \left(x_{i,1}, \cdots, x_{i,n_x}\right)^T \in R^{n_x}, y_i = \left(y_{i,1}, \cdots, y_{i,n_y}\right)^T \in R^{n_y}$。我们实现两层全连接网络如下

$$z_{1,i} = W^{(1)} x_i + b^{(1)} \in R^{n_1}, W^{(1)} \in R^{n_1 \times n_x}, b^{(1)} \in R^{n_1}$$

$$h_{1,i} = Sigmoid\left(z_{1,i}\right) \in R^{n_1}$$

$$z_{2,i} = W^{(2)} h_{1,i} + b^{(2)} \in R^{n_y}, W^{(2)} \in R^{n_y \times n_1}, b^{(2)} \in R^{n_y}$$

$$\hat{y}_i = Softmax\left(z_{2,i}\right)$$

$$f\left(\{x_i, y_i\}, W^{(1)}, b^{(1)}, W^{(2)}, b^{(2)}\right) = -\frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{n_y} y_{i,k} \log \hat{y}_{i,k}$$

试推导$\frac{\partial f}{\partial W^{(1)}}, \frac{\partial f}{\partial b^{(1)}}, \frac{\partial f}{\partial W^{(2)}}, \frac{\partial f}{\partial b^{(2)}}$。

提示：Softmax函数是Sigmoid函数的多值版本

典型例题

- 反向传播算法推导
  - 利用链式法则，对第i个样本有

$$\frac{\partial f}{\partial \hat{y}_{i,j}} = -\frac{1}{n} y_{i,j} \cdot \frac{1}{\hat{y}_{i,j}}$$

$$\frac{\partial \hat{y}_{i,j}}{\partial z_{2,i,k}} = \begin{cases} \hat{y}_{i,j}(1-\hat{y}_{i,k}), j=k \\ -\hat{y}_{i,j} \cdot \hat{y}_{i,k}, j \neq k \end{cases}$$

$$\frac{\partial z_{2,i,k}}{\partial b_k^{(2)}} = 1, \frac{\partial z_{2,i,k}}{\partial W_{k,l}^{(2)}} = h_{1,i,l}$$

$$\frac{\partial h_{1,i,l}}{\partial z_{1,i,l}} = h_{1,i,l}(1-h_{1,i,l})$$

$$\frac{\partial z_{1,i,l}}{\partial b_l^{(1)}} = 1, \frac{\partial z_{1,i,l}}{\partial W_{l,s}^{(1)}} = x_{i,s}$$

$\hat{y}_i$的每个元素都与$z_{2,i}$的每个元素有关联！故

$$\frac{\partial f}{\partial z_{2,i,k}} = \sum_{j=1}^{n_y} \frac{\partial f}{\partial \hat{y}_{i,j}} \cdot \frac{\partial \hat{y}_{i,j}}{\partial z_{2,i,k}}$$

重要知识点

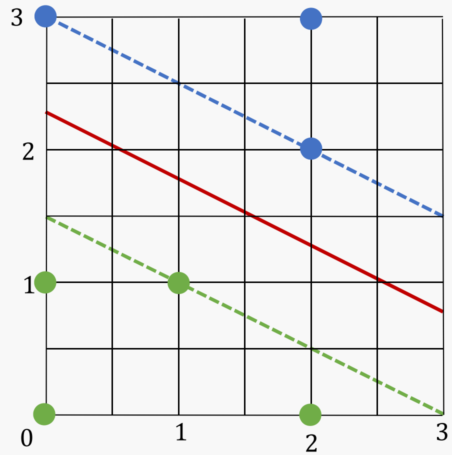
- 第五章 支持向量机
  - 间隔与支持向量
    - 了解支持向量机的物理含义
    - 能够利用几何性质求解简单的支持向量机模型
  - 对偶问题
    - 熟悉对偶问题的推导
    - 能利用对偶优化问题求解支持向量机
  - 核函数
    - 能进行基函数和核函数之间的转换与计算
    - 能够求解使用了核函数的支持向量机
  - 软间隔支持向量机、支持向量回归
    - 了解支持向量机的各种形式变体，能够根据图形对号入座

典型例题

利用线性SVM求解分类问题

$$\boldsymbol{\omega_1}: \{(\mathbf{1},\mathbf{1})^T,(\mathbf{0},\mathbf{1})^T,(\mathbf{2},\mathbf{0})^T,(\mathbf{0},\mathbf{0})^T\}$$

$$\boldsymbol{\omega_2}: \{(\mathbf{2},\mathbf{2})^T,(\mathbf{2},\mathbf{3})^T,(\mathbf{0},\mathbf{3})^T\}$$



解答：通过观察法获取支持向量为

$$\omega_1: (1,1)^T \quad \omega_2: (2,2)^T,(0,3)^T$$

列式计算

$$w^T\begin{pmatrix}1\\1\end{pmatrix}+b=1 \quad w^T\begin{pmatrix}0\\3\end{pmatrix}+b=-1 \quad w^T\begin{pmatrix}2\\2\end{pmatrix}+b=-1$$

解得

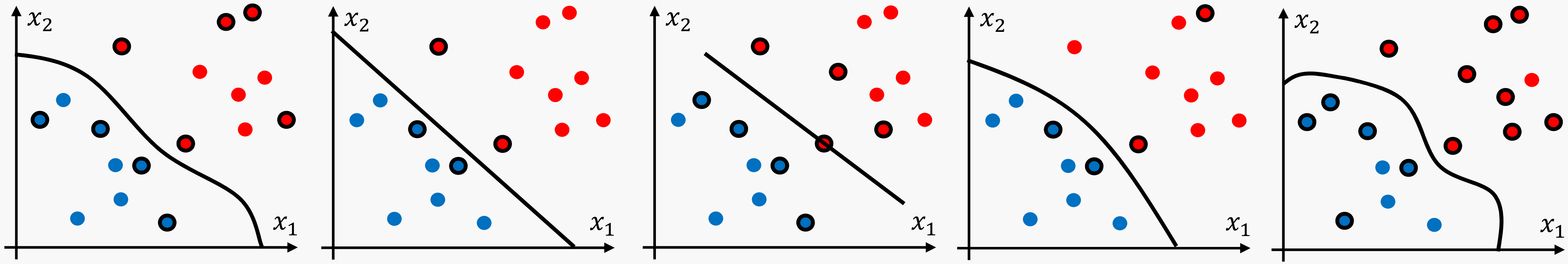
$$w=\begin{pmatrix}-\frac{4}{3}\\-\frac{2}{3}\end{pmatrix}, \qquad b=3, \qquad -\frac{4}{3}x_1-\frac{2}{3}x_2+3=0$$

典型例题

- 为下面的每个SVM模型标出正确的图形

$$\min \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{N} \xi_i$$
$$\text{s.t.}\ \xi_i \geq 0, y_i(\boldsymbol{w}^T\boldsymbol{x} + w_0) \geq 1 - \xi_i, C = 0.1$$

$$\min \frac{1}{2}||\boldsymbol{w}||^2 + C\sum_{i=1}^{N} \xi_i$$
$$\text{s.t.}\ \xi_i \geq 0, y_i(\boldsymbol{w}^T\boldsymbol{x} + w_0) \geq 1 - \xi_i, C = 1$$

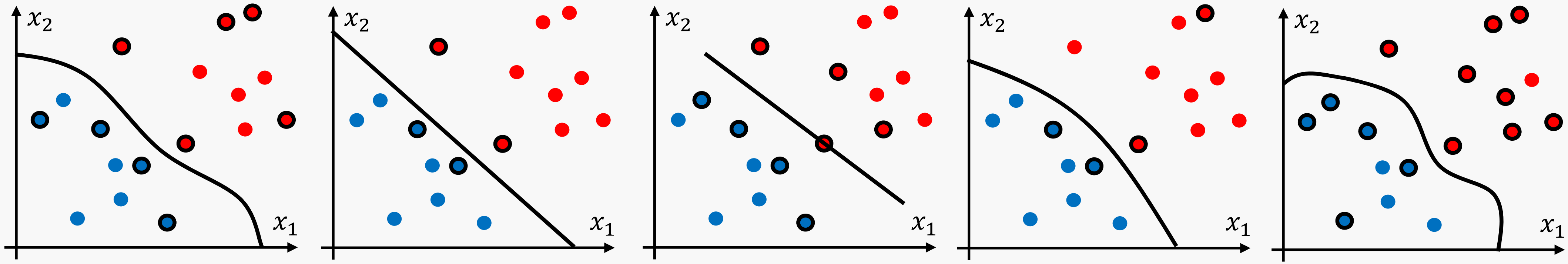


典型例题

- 为下面的每个SVM模型标出正确的图形

$$\max\left(\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j k(\boldsymbol{x}_i,\boldsymbol{x}_j)\right)$$

$$\text{s.t. } \alpha_i\geq 0,\sum_{i=1}^{N}\alpha_i y_i=0, k(\boldsymbol{x},\boldsymbol{x}')=\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}'+\left(\boldsymbol{x}^{\mathrm{T}}\boldsymbol{x}'\right)^2$$



典型例题

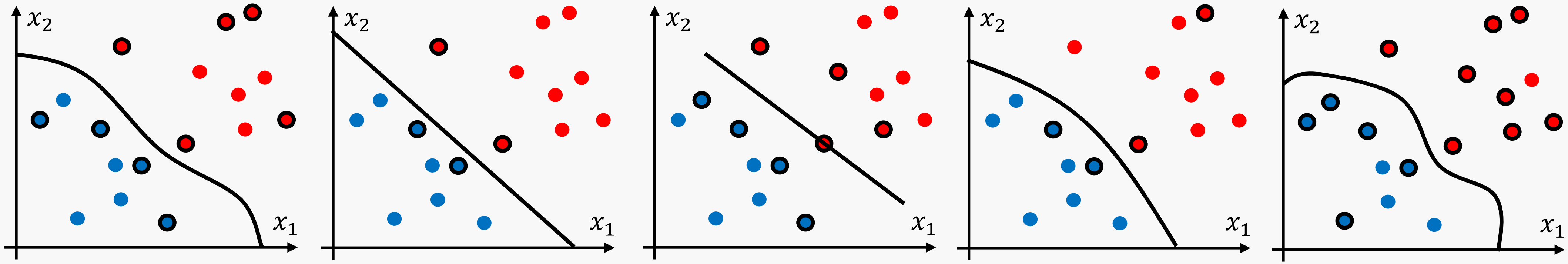
- 为下面的每个SVM模型标出正确的图形

$$\max\left(\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j k(\boldsymbol{x}_i,\boldsymbol{x}_j)\right)$$

$$\mathrm{s.t.}\ \alpha_i\geq 0,\sum_{i=1}^{N}\alpha_i y_i=0, k(\boldsymbol{x},\boldsymbol{x}')=\exp\left(-\frac{1}{2}||\boldsymbol{x}-\boldsymbol{x}'||^2\right)$$

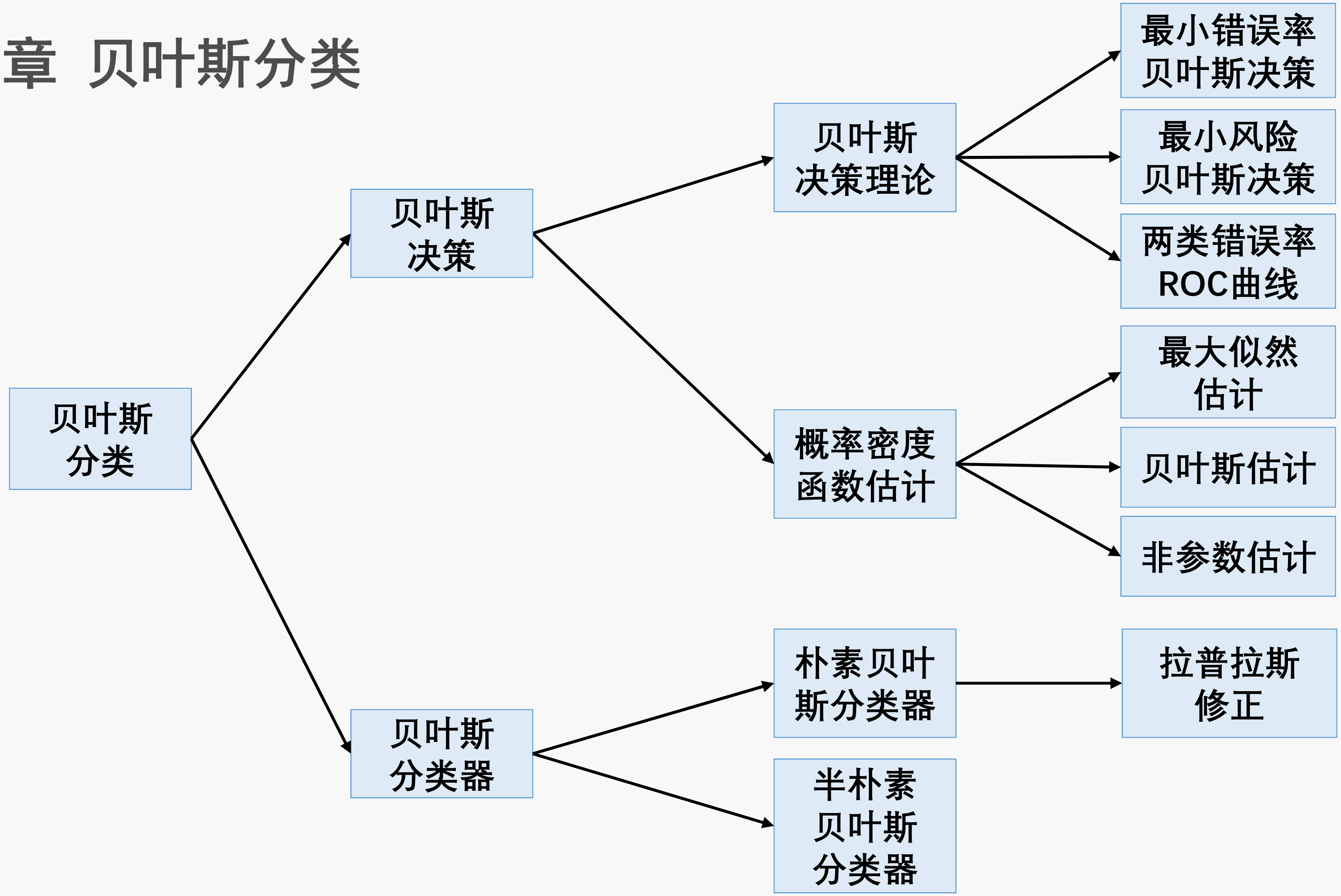
$$\max\left(\sum_{i=1}^{N}\alpha_i-\frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j k(\boldsymbol{x}_i,\boldsymbol{x}_j)\right)$$

$$\mathrm{s.t.}\ \alpha_i\geq 0,\sum_{i=1}^{N}\alpha_i y_i=0, k(\boldsymbol{x},\boldsymbol{x}')=\exp\left(-||\boldsymbol{x}-\boldsymbol{x}'||^2\right)$$



复习举例

第六章 贝叶斯分类



重要知识点

- 第六章 贝叶斯分类
  - 贝叶斯决策理论基本概念
    - 最小错误率贝叶斯决策与最小风险贝叶斯决策的区分
    - 两类错误率、ROC曲线基本概念
  - 概率密度函数估计计算
    - 最大似然估计计算
    - 贝叶斯估计计算
    - 非参数估计基本概念
  - 朴素贝叶斯分类器基本概念

典型例题

- 假定进行癌细胞识别，已知：
  - 正常细胞$P(\omega_1)=0.9$；癌细胞$P(\omega_2)=0.1$
  - 若观察结果为x，从类条件概率密度曲线上查得，$P(x \mid \omega_1)=0.2$；$P(x \mid \omega_2)=0.4$
  - $\lambda(a_1,w_2)=6$；$\lambda(a_2,w_1)=1$；其余为0
- 如果根据最小错误率准则：

$$P(x \mid \omega_1)P(\omega_1)=0.2*0.9=0.18$$

$$P(x \mid \omega_2)P(\omega_2)=0.4*0.1=0.04$$

  - 决策为第一类

典型例题

- 假定进行癌细胞识别，已知：
  - 正常细胞$P(\omega_1) = 0.9$；癌细胞$P(\omega_2) = 0.1$
  - 若观察结果为x，从类条件概率密度曲线上查得，$P(x \mid \omega_1) = 0.2$；$P(x \mid \omega_2) = 0.4$
  - $\lambda(a_1, w_2) = 6$；$\lambda(a_2, w_1) = 1$；其余为0
- 如果根据最小风险准则：

$$R(\alpha_i \mid x) = \sum_{j=1}^{2} \lambda(\alpha_i \mid \omega_j) P(\omega_j \mid x)$$

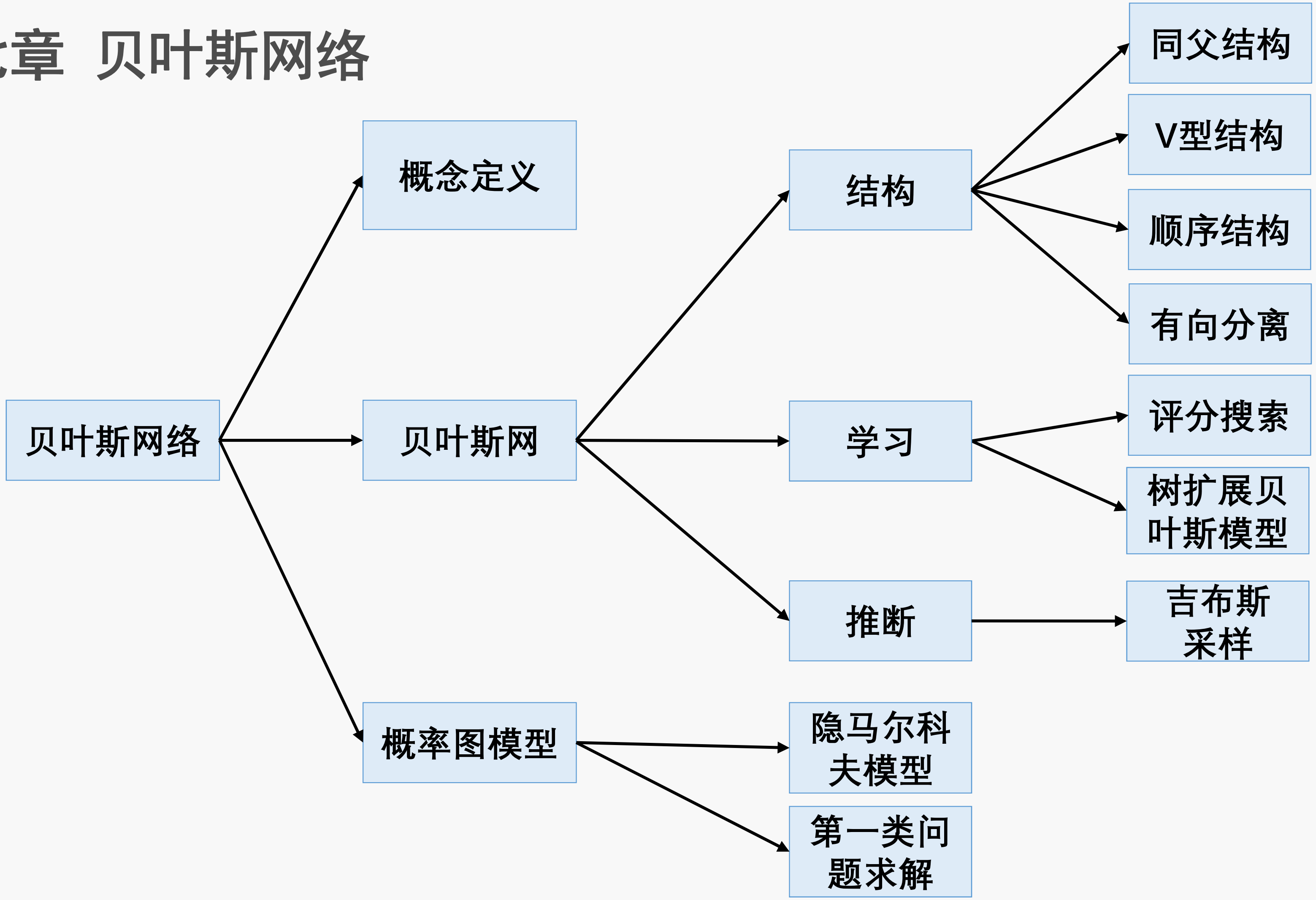
有 $R(\alpha_1 \mid x) = \lambda(\alpha_1 \mid \omega_2) P(\omega_2 \mid x) = 1.092$

$R(\alpha_2 \mid x) = \lambda(\alpha_2 \mid \omega_1) P(\omega_1 \mid x) = 0.818$

  - 决策为第二类

复习举例

□ 第七章 贝叶斯网络

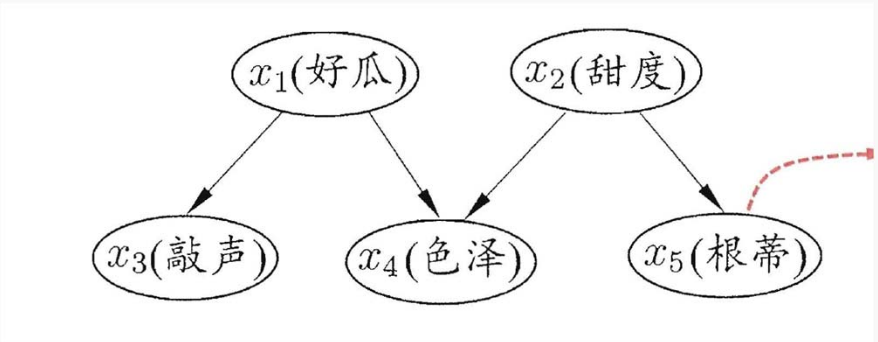


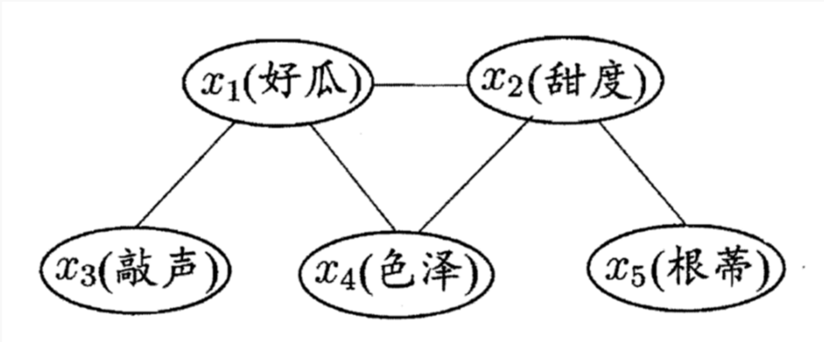
重要知识点

- 第七章 贝叶斯网络
  - 贝叶斯网络的基本概念
    - 贝叶斯网的结构
    - 贝叶斯网的学习
    - 贝叶斯网的推断
  - 贝叶斯网的有向分离

典型例题

- 左图的道德图如右图所示
  - 从图中能容易地找出所有的条件独立关系

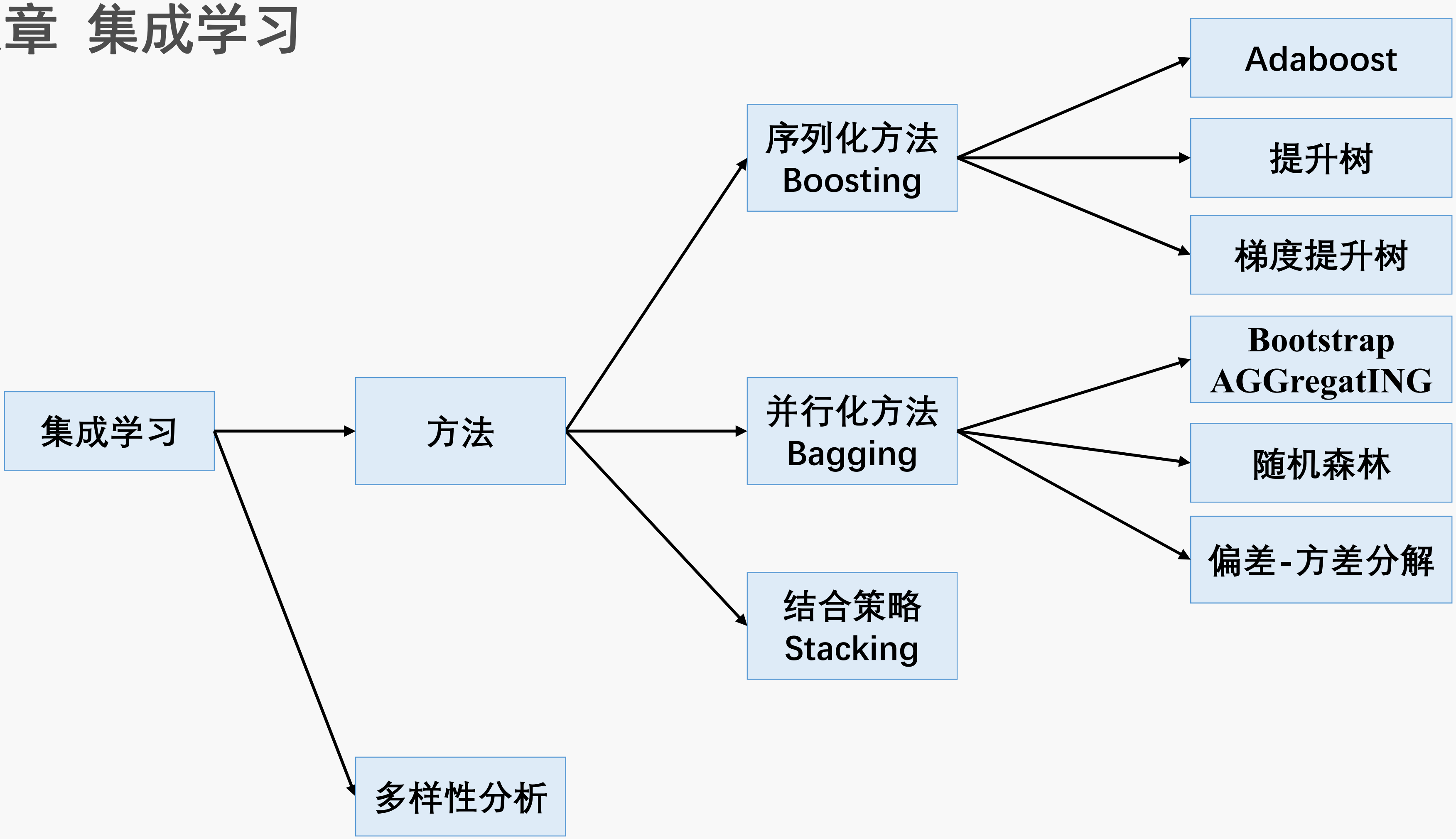




$x_3 \perp x_4 | x_1$，$x_4 \perp x_5 | x_2$，$x_3 \perp x_2 | x_1$，$x_3 \perp x_5 | x_1$，$x_3 \perp x_5 | x_2$

复习举例

第八章 集成学习



重要知识点

- 第八章 集成学习
  - 三种集成学习方法的基本概念
    - 序列化方法 boosting
    - 并行化方法bagging
    - 结合策略stacking

典型例题

Adaboost

- 训练数据如下表所示，假设弱分类器由$x < v$或$x > v$产生，其阈值v使该分类器在训练数据集上分类误差率最低。用AdaBoost算法学习一个强分类器。

序号	1	2	3	4	5	6	7	8	9	10
x	0	1	2	3	4	5	6	7	8	9
y	1	1	1	−1	−1	−1	1	1	1	−1

- **解答：** 首先初始化数据权值分布

$$\mathbb{D}_1 = (0.1, 0.1, \dots, 0.1)$$

- <u>对$t = 1$</u>
- 在权值分布为$\mathbb{D}_1$的训练数据上，阈值v取2.5时分类误差率最低，故基本分类器为

$$h_1(x) = \begin{cases} 1, x < 2.5 \\ -1, x > 2.5 \end{cases}$$

清華大學

典型例题

Adaboost

- $h_1(x)$在训练数据集上的误差率$\epsilon_1 = 0.3$
- 计算$h_1(x)$的系数：

$$\alpha_1 = \frac{1}{2}\ln\left(\frac{1-\epsilon_1}{\epsilon_1}\right) = 0.4236$$

- 更新训练数据的权值分布：

$$\mathrm{D}_2(x) = \frac{\mathrm{D}_1(x)e^{-f(x)\alpha_1 h_1(x)}}{Z_1}$$

- $\mathrm{D}_2 = (0.07143, 0.07143, 0.07143, 0.07143, 0.07143, 0.07143, 0.16667, 0.16667, 0.16667, 0.07143)$
- 组合分类器为$H(x) = \mathrm{sign}\big(0.4236h_1(x)\big)$，在训练集上有3个误分类点
- <u>对$t = 2$</u>
- 在权值分布为D_2的训练数据上，阈值v取8.5时分类误差率最低

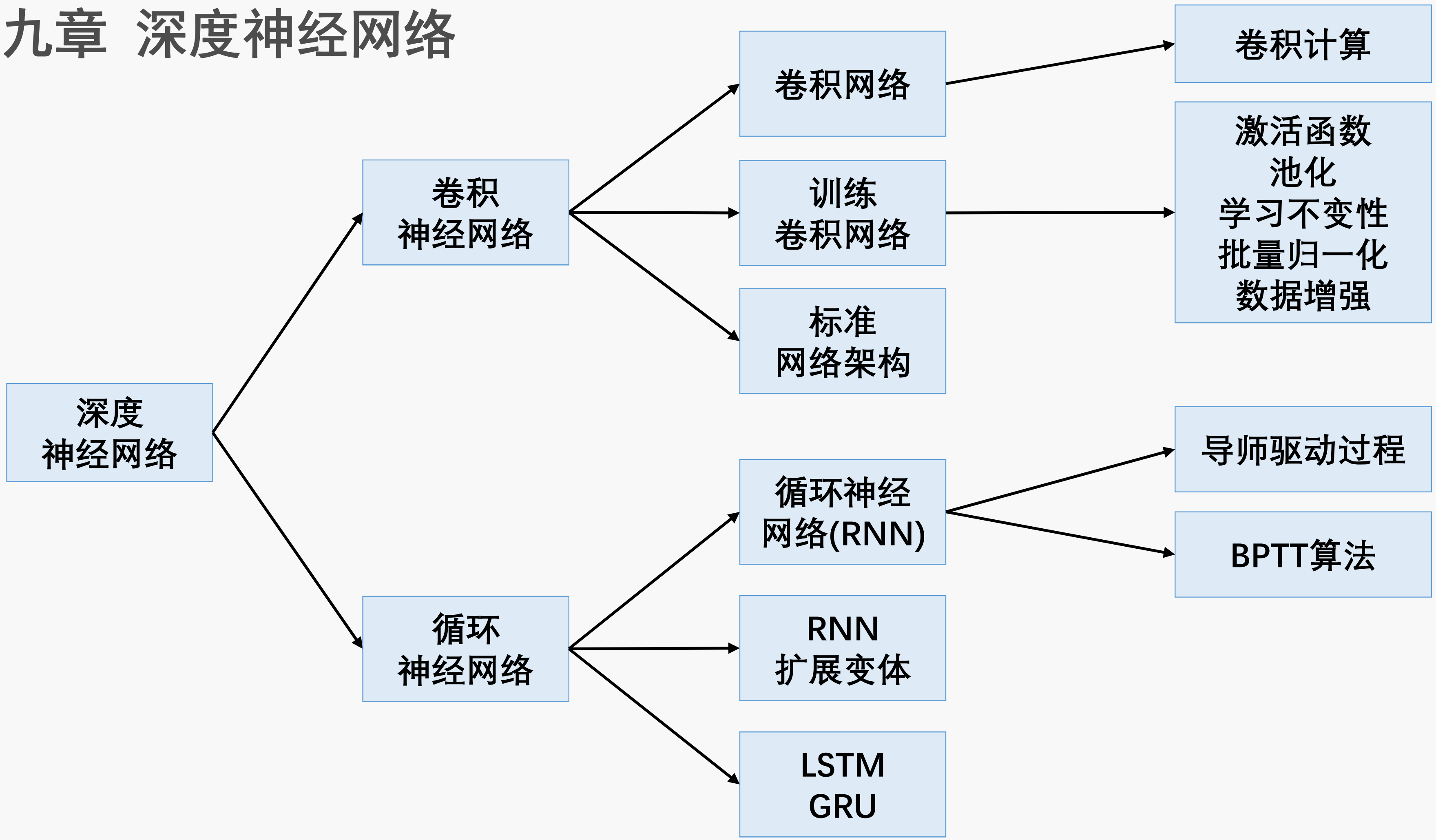
典型例题

Adaboost

- 在权值分布为$\mathbb{D}_2$的训练数据上，阈值v取8.5时分类误差率最低，得到$h_2(x)$
- 中间过程跳过，得到$\alpha_2 = 0.6496$和新的数据分布$\mathbb{D}_3$
- 组合分类器为$H(x) = \mathrm{sign}\big(0.4236h_1(x) + 0.6496h_2(x)\big)$，在训练集上有3个误分类点
- <u>对$t = 3$</u>
- 过程省略
- 组合分类器为$H(x) = \mathrm{sign}\big(0.4236h_1(x) + 0.6496h_2(x) + 0.7514h_3(x)\big)$，在训练集上有0个误分类点
- 以此作为最终分类器

复习举例

◻ 第九章 深度神经网络

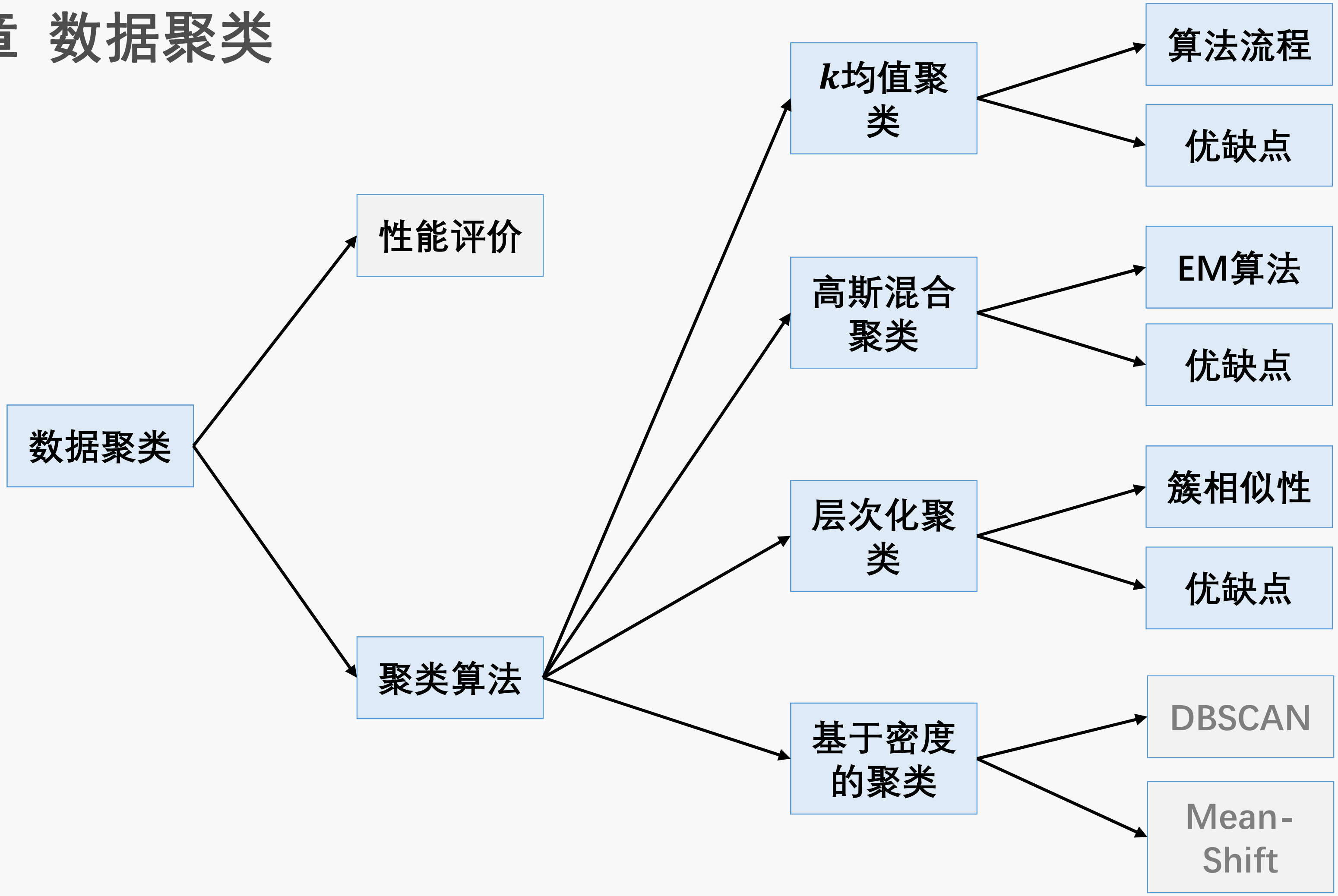


重要知识点

- 第九章 深度神经网络
  - 基本概念、定义的理解与掌握
    - 卷积的计算
    - 卷积网络训练技巧

复习举例

第十章 数据聚类



重要知识点

- 第十章 数据聚类
  - 给定少量数据，使用不同的聚类算法进行简单计算
  - 掌握不同聚类算法的特点，能够对比它们的优缺点
  - 了解影响聚类性能的因素

典型例题

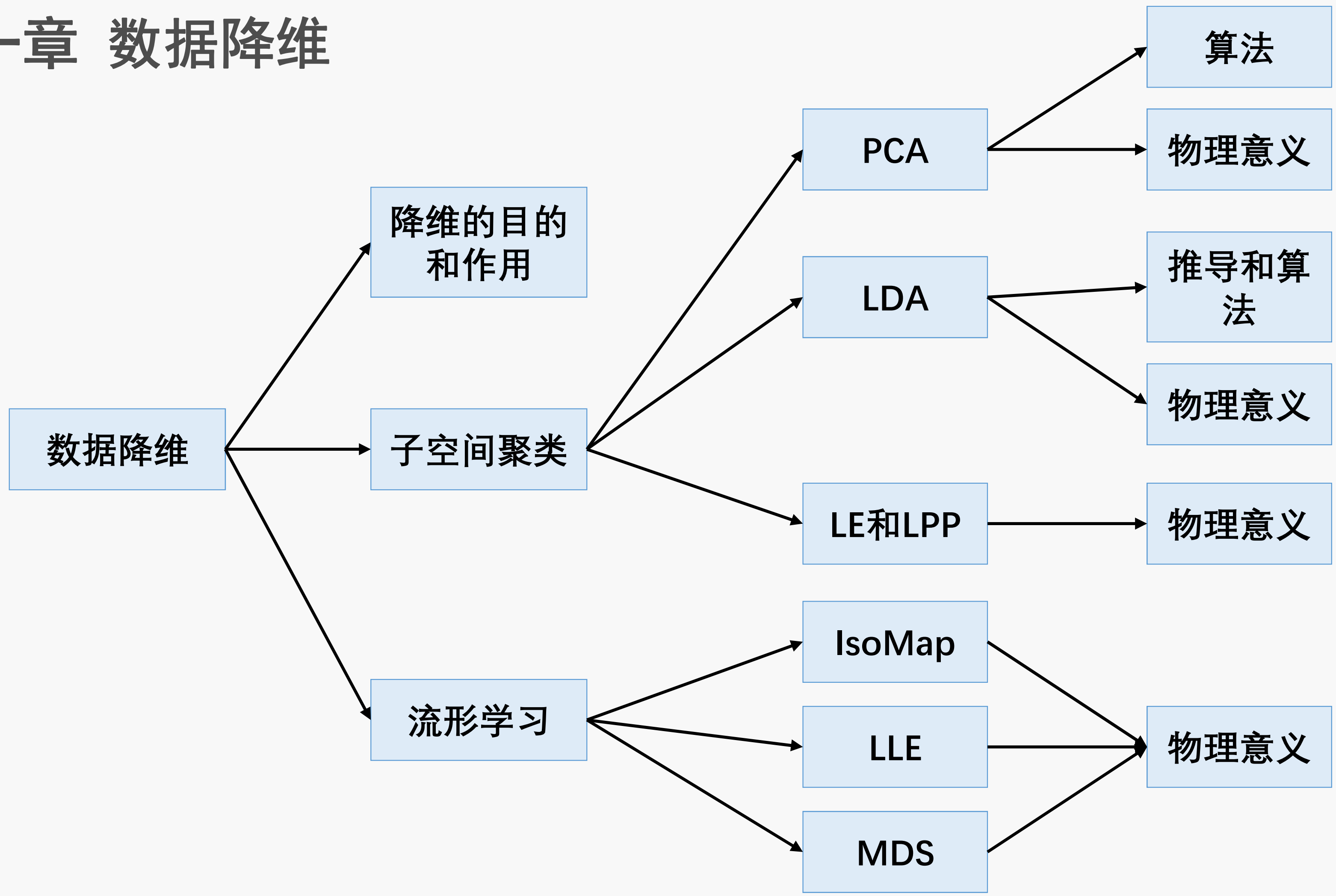
- 假设我们的样本集$D=\{\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_n\}$来自于$k$个高斯分布，即$p_M(\boldsymbol{x})=\sum_{j=1}^{k}\alpha_j\mathcal{N}\left(\boldsymbol{x}\middle|\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j\right)$，其中$\alpha_j,\boldsymbol{\mu}_j,\boldsymbol{\Sigma}_j$是混合高斯分布第$j$个成分的比例、均值以及协方差矩阵，试推导采用EM算法求解混合高斯模型参数的过程。

  （1）根据贝叶斯公式，求后验期望，写出需要最大化的函数形式。

  （2）最大化（1）中的函数，写出模型参数α_j,μ_j,Σ_j的更新公式。

复习举例

第十一章 数据降维



重要知识点

- 第十一章 数据降维
  - 数据降维的目的和作用
  - 推导PCA和LDA及简单的变体
  - 熟练运用PCA和LDA进行数据降维
  - 了解子空间降维和流形学习的概念，了解其物理意义

典型例题

我们曾经学习过Fisher判别方法，它将两类样本投影到类间方差尽可能大而类内方差尽可能小的方向上。这里我们考虑将Fisher判别进行推广。

考虑K分类问题。我们尝试利用某种线性变换$W \in \mathbb{R}^{D \times D'}$，将原始的$D$ $(D > K)$维样本$x \in \mathbb{R}^{D}$投影到D' $(D' < D)$维的空间中，投影结果记为$y \in \mathbb{R}^{D'}$，则x与y的关系可表示为：

$$y = W^T x$$

通过以上投影方式，我们便能将原始D维的特征降维到D'维空间中，实现特征降维。重新定义K分类问题的类内离散度矩阵与类间离散度矩阵：

$$S_W = \sum_{k=1}^{K} \sum_{n \in C_k} (x_n - m_k)(x_n - m_k)^T ; S_B = \sum_{k=1}^{K} N_k (m_k - m)(m_k - m)^T$$

其中，$m_k = \frac{1}{N_k} \sum_{n \in C_k} x_n$，$N_k$表示第$k$类样本的总数。

典型例题

（1）请证明类内离散度矩阵S_W与类间离散度矩阵S_B的和等于总离散度矩阵，即：

$$S_W + S_B = S_T = \sum_{n=1}^{N} (x_n - m)(x_n - m)^T$$

其中，$m = \frac{1}{N}\sum_{n=1}^{N} x_n = \frac{1}{N}\sum_{k=1}^{K} N_k m_k$，$N$为样本总数。

（2）请写出投影之后的样本在D'维空间中的类内离散度矩阵$\tilde{S}_W$与类间离散度矩阵$\tilde{S}_B$的表达式。

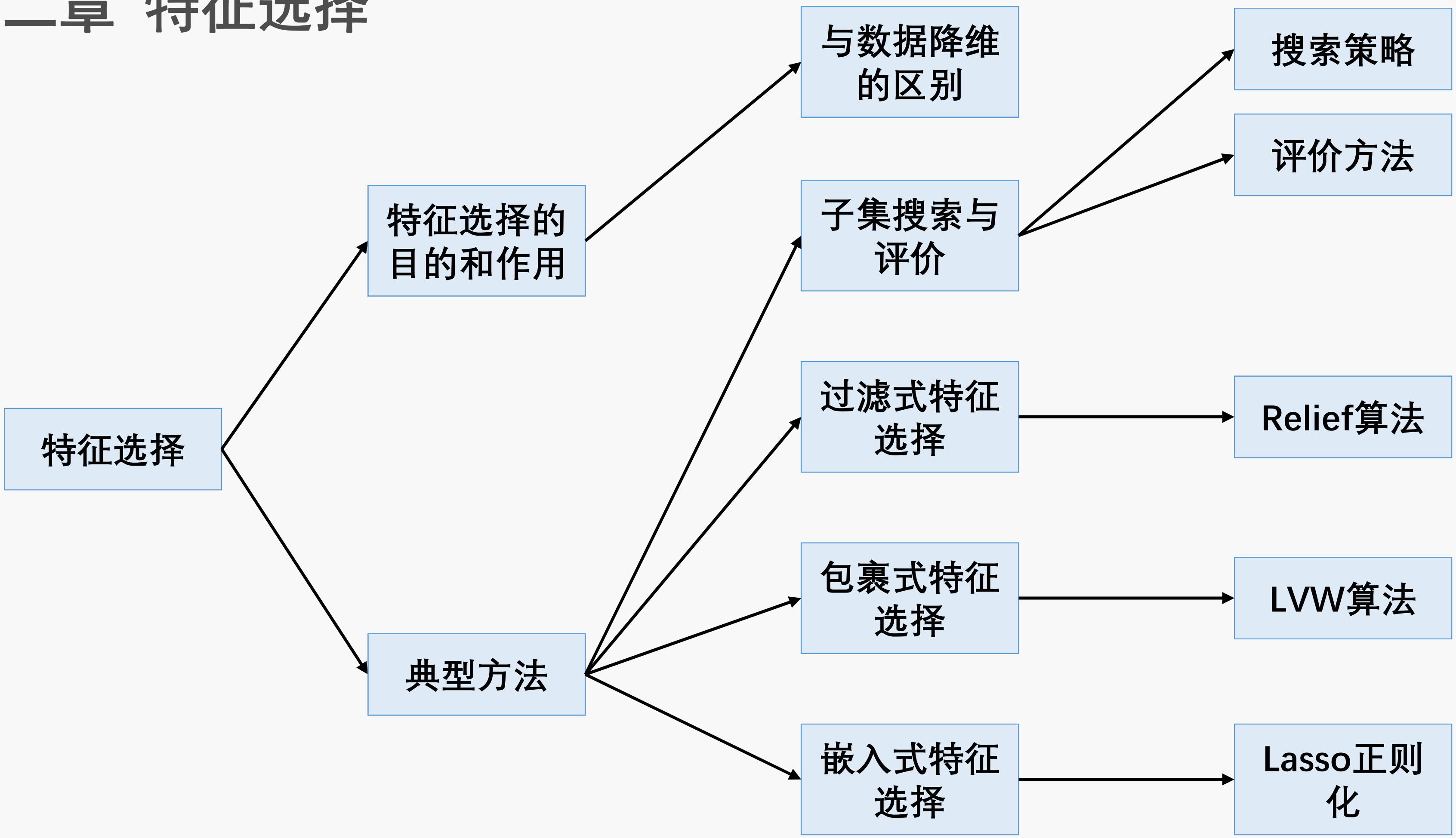
（3）类似投影到一维的Fisher判别，我们用如下判据来确定变换矩阵W：

$$J(W) = \frac{\prod_{diag} \tilde{S}_B}{\prod_{diag} \tilde{S}_W}$$

其中，$\prod_{diag} A$表示矩阵A的对角线元素的连乘。请用W, S_W, S_B表示出$J(W)$，W的每一列记为$w_k \in \mathbb{R}^D, k = 1, 2, \dots, D'$。并最大化$J(W)$，求解变换矩阵$W$。

复习举例

第十二章 特征选择



重要知识点

- 第十二章 特征选择
  - 特征选择的目的及作用，与特征降维的区别
  - 前向、后向子集搜索算法，主要的子集评价策略
  - 过滤式、包裹式、嵌入式特征选择方法的思想和区别

典型例题

- 以下描述中属于特征选择情况是（）

（A）提取向量偶数位特征，重新表示向量

（B）前向序贯搜索子集，进行子集评价，选择最优子集

（C）采用可分性度量，度量每个特征，进行选择

（D）主成分分析进行特征降维

重要知识点

- 第十三章 度量学习
  - 度量空间
    - 利用度量空间对距离进行判断和证明
    - 了解度量矩阵的定义与计算
  - 度量学习方法
    - 对各种距离度量形式有了解，根据公式定义推导出对应性质
    - 对线性度量学习（如LMNN）的经典方法有基本了解
    - 对深度度量学习的三元组损失函数有基本了解

重要知识点

- 第十四章 半监督学习
  - 了解半监督学习的定义、未标记样本的性质
  - 对经典的半监督学习算法有基本了解
  - 半监督学习算法常遵循的假设

期末考试安排

- **考试时间**：2023年1月20日，周二晚上14:30-16:30
- **考试地点**：舜德楼/经管西楼401
- **考试形式**：开卷考试，但只准带课本和计算器
- **考试题型**：判断题、选择题、填空题、计算题
- **答疑时间**：1月19日，8:30-11:30和14:00-17:00
- **答疑地点**：中央主楼623会议室
- **考试证件**：学生证

预祝大家取得优异成绩！