Machine Learning and Pattern Recognition

东北大学“智能+X”新工科课程系列

机器学习与模式识别

东北大学 信息科学与工程学院
人工智能系 、智能感知与机器人研究所
陈东岳

Exercises Explanation

习题讲解

CHAPTER ONE

第一章 概述

Section 1, Introduction

第一章 概述

判断题

- 模型容量大小与训练集样本数量无关。√
  - 解析：模型容量与模型的数学形式有关，表现为假设空间大小
- 班级内自由分组讨论不属于分类问题。√
  - 解析：“自由分组”意味着没有任务定义的“类”的概念。
- 驯兽师用鞭打和喂食的方式教导老虎跳火圈属于有监督学习。×
  - 解析：没有具体如何跳的“答案”，只有奖惩信号，是强化学习

第一章 概述

选择题

- 从当前考场的所有考生中，找出作弊的学生，该任务属于

A. 分类	B. 聚类	C. 异常检测	D．以上均是

- 解析：
- 1. 存在任务定义的“正常”与“异常”两类，正常类占明显多数，具有相对统一的模式。
- 2. 不是聚类，“正常”与“异常”存在未知的标准答案，不是根据数据分布自主产生的。此外，异常类不一定具有明显的聚类性，

▶ 简答题

- 人脸图像伪造属于分类任务还是回归任务，为什么？
  - 答：人脸伪造图像属于回归任务。原因是人脸伪造算法生成的图像可以看作是一个高维连续值变量，且没有明确的类别信息。
- 如果一个模型在训练集上表现良好，但在测试集上表现不佳，试说明该模型的容量与问题复杂度之间的关系？
  - 答：该模型的容量可能明显超出了问题复杂度，导致了过拟合现象

第一章 概述

计算题

- 下面是一个房屋价格预测数据集的部分数据，该任务是分类任务还是回归任务？是有监督还是无监督？试写出每个样本的特征向量？特征空间的维度是多少？

编号	占地面积/m^2	距市中心距离/km	地下室面积/ m^2	房屋价格/万元
1	100	12	3	81.5
2	120	5	6	126
3	90	18	5	76.2
4	150	3	10	165.3
5	200	20	0	161
6	160	9	8	152.6

- 解：该任务是回归任务，属于有监督学习；1号样本的样本特征向量为$[100,12,3]^T$,依此类推可以写出每个样本的特征向量；特征空间维度为3.

CHAPTER TWO

第二章 贝叶斯决策论

Section 2, Bayesian Decision

第二章 贝叶斯决策论

填空题

- 面向d维数据的线性分类器决策面是d维空间中的一个$(d-1)$维的（超平面）。
  - 解析：对比二维空间上的线性决策面是1条直线，三维空间中的线性决策面是一个二维平面。

判断题

- 在分类问题中，类条件概率是概率质量，后验概率是概率密度。 ×
  - 解析：类条件概率是关于连续变量x的概率，是概率密度。
  - 解析：后验概率是关于离散类别标签ω的概率，是概率质量。
- 有些问题中，无需了解先验概率也可以直接估计后验概率。 √
  - 解析：判别模型可以直接估计后验概率，例如逻辑回归

选择题

- 假设样本$\boldsymbol{x} \in \Re^3$，服从多元正态分布$\mathcal{N}(\boldsymbol{\mu}, \Sigma)$，则类条件概率密度函数的有效参数数量为：

A. 12个　　B. 9个　　C. 6个　　D．不确定

- 解析：
- 均值向量$\boldsymbol{\mu} \in \Re^3$, 包含3个有效参数
- 协方差矩阵$\Sigma \in \Re^{3\times3}$, 由于协方差矩阵为对称阵，因此6个非对角线元素上只有3个有效参数，加上对角线元素的3个有效参数，共6个有效参数
- 结论：共9个有效参数。

第二章 贝叶斯决策论

▶ 简答题

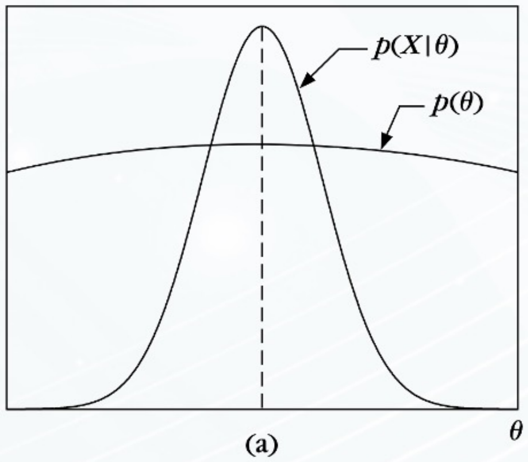
- 如果对于某个数据集的概率分布模型，最大似然估计和最大后验概率估计给出结果完全相同，试解释可能造成这一结果的原因（可画图说明）
- 答：

最大似然估计的结果是似然函数的最大值解，记为$\underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, p(X;\boldsymbol{\theta})$；

最大后验估计的结果是后验概率的最大值解，记为$\underset{\boldsymbol{\theta}}{\operatorname{argmin}}\, p(X;\boldsymbol{\theta})\, p(\boldsymbol{\theta})$

如果两种方法的解完全相同，

说明似然函数$p(X;\boldsymbol{\theta})$与参数$\boldsymbol{\theta}$的先验概率分布可能存在相同的最大解。



MAP与MLE结果相同的情况

▶ 计算题

- 已知两个类别ω_1和ω_2，其先验概率相等，两类的类条件概率服从正态分布，有$p(\boldsymbol{x}|w_1)=\mathcal{N}(\mu_1,\Sigma)$和$p(\boldsymbol{x}|w_2)=\mathcal{N}(\mu_2,\Sigma)$，且$\mu_1=\begin{bmatrix}0\\0\end{bmatrix}$，$\mu_2=\begin{bmatrix}3\\3\end{bmatrix}$，$\Sigma=\begin{bmatrix}1.1 & 0.3\\0.3 & 1.9\end{bmatrix}$，试采用贝叶斯决策对样本$\boldsymbol{x}=\begin{bmatrix}1.0\\2.2\end{bmatrix}$进行分类。

解：

根据贝叶斯定理：$P(\omega|\boldsymbol{x})=\frac{p(\boldsymbol{x}|\omega)P(\omega)}{p(\boldsymbol{x})}$

其中的$p(\boldsymbol{x})$对每个类别相同，且两类先验概率$P(\omega)$相等，因此判别函数可以写为：

$$g_i(x)=\ p(\boldsymbol{x}|\omega_i)$$

代入相关参数与样本：

$$g_1(x)=\exp\left(-\frac{1}{2}\begin{bmatrix}1.0-0\\2.2-0\end{bmatrix}^T\begin{bmatrix}1.1 & 0.3\\0.3 & 1.9\end{bmatrix}^{-1}\begin{bmatrix}1.0-0\\2.2-0\end{bmatrix}\right)=e^{-1.476}=\ 0.159$$

$$g_2(x)=\exp\left(-\frac{1}{2}\begin{bmatrix}1.0-3\\2.2-3\end{bmatrix}^T\begin{bmatrix}1.1 & 0.3\\0.3 & 1.9\end{bmatrix}^{-1}\begin{bmatrix}1.0-3\\2.2-3\end{bmatrix}\right)=e^{-1.836}=\ 0.229$$

由于$g_2(x)>g_1(x)$，因此样本$\boldsymbol{x}$属ω_2类别

计算题

- 已知iris数据中的setosa类的10个样本的部分特征数据如下，利用朴素贝叶斯分类器+直方图法(bin边长为0.4cm)估计花萼长度=4.8cm, 花萼宽度=3.2cm的样本的类条件概率密度.
- 解：首先利用直方图法，
- 1. 花萼长度范围为4.4~5.4，以4.4为第一个bin的中心，bin宽度为0.4, bin空间前闭后开, 则花萼长度4.8cm所处的bin为$B_{1,2}=[4.6,5.0)$，包含5个样本，该bin的概率密度函数：

$$P(x_1=4.8|\text{setosa})=\frac{5}{10}\times\frac{1}{0.4}=\frac{5}{4}$$

- 2.花萼宽度范围为2.9~3.9，以2.9为第一个bin的中心，bin宽度为0.4, bin空间前闭后开，则花萼宽度3.2cm所处的bin为$B_{2,2}=[3.1,3.5)$，包含5个样本，该bin的概率密度函数：

$$P(x_2=3.2|\text{setosa})=\frac{5}{10}\times\frac{1}{0.4}=\frac{5}{4}$$

- 根据朴素贝叶斯分类器：

$$p(\boldsymbol{x}|\omega)=\prod_{i=1}^{d=2}P(x_i|\omega)=\frac{5}{4}\times\frac{5}{4}=1.5625$$

序号	花萼长度/cm	花萼宽度/cm	类别
1	5.1	3.5	setosa
2	4.9	3	setosa
3	4.7	3.2	setosa
4	4.6	3.1	setosa
5	5	3.6	setosa
6	5.4	3.9	setosa
7	4.6	3.4	setosa
8	5	3.4	setosa
9	4.4	2.9	setosa
10	4.9	3.1	setosa

CHAPTER THREE

第三章 线性模型

Section 3, Linear Model

第三章 线性模型

▶ 填空题

- 线性回归模型的封闭解又称为（ 最小二乘 ）解。
  - 解析：由于线性回归模型使用了均方误差目标函数，因此其封闭解是该问题的最小二乘解。
- 逻辑回归函数的取值区间是（ (0, 1) ）。
  - 解析：logistic函数无法取到1或0，因此是开区间。

▶ 判断题

- 线性回归模型如存在唯一解，必然可以令模型在训练集上的均方误差为0. ×
  - 解析：唯一解表示均方误差取最小值，不一定是0
- 逻辑回归模型无法用于多类分类问题。 ×
  - 解析：将多类问题分解为多个二分类问题，就可以用逻辑回归求解

▶ 选择题

- 假设线性回归问题的训练集为：$\boldsymbol{x}_i \in \Re^3, i = 1, \ldots, 100$，则该线性模型包含多少个参数：

A. 3个　　B. 4个　　C. 100个　　D . 101

  - 解析：$y = w_1x_1 + w_2x_2 + w_3x_3 + w_0$, 参数$w_j, j = 0,1,2,3$共4个参数

- 逻辑回归模型中的逻辑回归函数可以看作是对以下哪种概率的描述：

A. $p(x|\omega_1)$　　B. $p(\omega_1|x)$　　C. $p(x)$　　D. $p(\omega_1)$

  - 解析：逻辑回归函数是在正态分布假设下对后验概率的估计。

▶ 简答题

- 在线性回归模型中，假设输入样本记为$\boldsymbol{x}_i, i = 1, \dots, N$，相应的类别标签记为$y_i, i = 1, \dots, N$。请给出自相关矩阵和互相关向量的定义（公式与符号表达）。如果样本集$X \in \Re^{N\times(d+1)}$中，$N < d + 1$，应如何处理才能得到合理的模型参数向量唯一解。

- 自相关矩阵$R_x = X^T X$，$X = \begin{bmatrix} \hat{\boldsymbol{x}}_1^T \\ \hat{\boldsymbol{x}}_2^T \\ \vdots \\ \hat{\boldsymbol{x}}_N^T \end{bmatrix}$；互相关向量记作$X^T\boldsymbol{y}$，其中$\boldsymbol{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$

- 当$N < d + 1$时，可以通过增加正则化项——$\lambda \|w\|^2$，得到新的最小二乘解如下：

$$\hat{\boldsymbol{w}} = (X^T X + \lambda I)^{-1} X^T \boldsymbol{y}$$

第三章 线性模型

简答题

- 标签y是一个随机变量，由函数$\hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}$加上一个随机噪声生成：$y=\hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}+\epsilon$；其中噪声$\epsilon$服从正态分布$\mathcal{N}(0,\sigma^2)$，设当前训练样本集为$X=\{\hat{\boldsymbol{x}}_i|i=1,\ldots,N\},Y=\{y_i|i=1,\ldots,N\}$，试推导$\hat{\boldsymbol{w}}$的最大似然解的数学形式；

答：根据定义，有$\epsilon=y-\hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}\sim\mathcal{N}(0,\sigma^2)$，其似然性函数为：$p(\boldsymbol{x};\delta^2,\hat{\boldsymbol{w}})=\frac{1}{\sqrt{2\pi\delta^2}}\exp\left(\frac{\left(y-\hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}\right)^2}{2\delta^2}\right)$

对于当前数据集X的整体对数似然性函数可以写为：

$$LL(X;\delta^2,\hat{\boldsymbol{w}})=\sum_{i=1}^N\ln\left(\frac{1}{\sqrt{2\pi\delta^2}}\exp\left(\frac{\left(y_i-\hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_i\right)^2}{2\delta^2}\right)\right)=\sum_i^N\left(\frac{\left(y-\hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_i\right)^2}{2\delta^2}-\frac{1}{2}\ln 2\pi-\frac{1}{2}\ln\delta^2\right)$$

则线性回归参数$\hat{\boldsymbol{w}}$的估计问题转化为以下优化问题：$\hat{\boldsymbol{w}}^*=\underset{\hat{\boldsymbol{w}}}{\operatorname{argmax}}\left(LL(X;\delta^2,\hat{\boldsymbol{w}})\right)$

令对数似然函数对线性回归函数的参数$\hat{\boldsymbol{w}}$求导为0，则有：

$$\frac{\partial LL(X;\delta^2,\hat{\boldsymbol{w}})}{\partial\hat{\boldsymbol{w}}}=\sum_{i=1}^N\left(\frac{(y_i-\hat{\boldsymbol{w}}^T\hat{\boldsymbol{x}}_i)\hat{\boldsymbol{x}}_i}{\delta^2}\right)=0\quad\Longrightarrow\quad\hat{\boldsymbol{w}}=(X^TX)^{-1}X^T\boldsymbol{y}$$

第三章 线性模型

▶ 简答题

- 如上题所述，假设$\widehat{\boldsymbol{w}}$的先验分布服从$d+1$元正态分布$\mathcal{N}\left(\mathbf{0}, I_{(d+1)\times(d+1)}\right)$,试采用最大后验概率估计法推导$\widehat{\boldsymbol{w}}$的最优解的数学形式

答：根据最大后验概率估计的定义，则目标函数可以写为：

$$J(X;\delta^2,\widehat{\boldsymbol{w}}) = \sum_{i=1}^{N} \ln\left(\frac{1}{\sqrt{2\pi\delta^2}}\exp\left(\frac{(y_i-\widehat{\boldsymbol{w}}^T\widehat{\boldsymbol{x}}_i)^2}{2\delta^2}\right)\right) + \ln\left(\frac{1}{\sqrt{2\pi}}\exp\left(\frac{\widehat{\boldsymbol{w}}^T\widehat{\boldsymbol{w}}}{2}\right)\right)$$

则目标函数导数为0的方程为：$\frac{\partial J(X;\delta^2,\widehat{\boldsymbol{w}})}{\partial\widehat{\boldsymbol{w}}} = \sum_{i=1}^{N}\left(\frac{(y_i-\widehat{\boldsymbol{w}}^T\widehat{\boldsymbol{x}}_i)\widehat{\boldsymbol{x}}_i}{\delta^2} + \widehat{\boldsymbol{w}}\right) = 0$

推导：$\sum_{i=1}^{N}\left((y_i - \widehat{\boldsymbol{w}}^T\widehat{\boldsymbol{x}}_i)\widehat{\boldsymbol{x}}_i + \delta^2\widehat{\boldsymbol{w}}\right) = 0 \Longrightarrow \left[\sum_{i=1}^{N}\left(\widehat{\boldsymbol{x}}_i\widehat{\boldsymbol{x}}_i^T + \delta^2 I\right)\right]\widehat{\boldsymbol{w}} = \sum_{i=1}^{N}\widehat{\boldsymbol{x}}_n y_n$

导出：$\widehat{\boldsymbol{w}} = (X^TX + N\delta^2 I)^{-1}X^T\boldsymbol{y}$

令$N\delta^2 = \lambda$, 则有$\widehat{\boldsymbol{w}} = (X^TX + \lambda I)^{-1}X^T\boldsymbol{y}$

计算题

- 已知3个样本坐标为$\boldsymbol{x}^{(1)}=[1,0],\boldsymbol{x}^{(2)}=[0,1],\boldsymbol{x}^{(3)}=[1,1]$，其中$\boldsymbol{x}^{(1)},\boldsymbol{x}^{(2)}\in\omega_1,\boldsymbol{x}^{(3)}\in\omega_2$,初始化逻辑回归模型对应的线性决策面为：$x_1=0.5$。设学习步长为0.2。请计算基于逻辑回归梯度下降法进行一次迭代后的决策面方程，并在下图中画出相应的决策面直线。

解：根据题意，定义以下变量并且初始化：步长：$\rho=0.2$；模型参数：$\hat{\boldsymbol{w}}=[w_1,w_2,w_0]=[1,0,-0.5]$；标签：$\boldsymbol{y}=[1,1,0]^T$

训练样本：$\hat{\boldsymbol{x}}^{(1)}=[1,0,1],\hat{\boldsymbol{x}}^{(2)}=[0,1,1],\hat{\boldsymbol{x}}^{(3)}=[1,1,1]$

计算梯度：

$$\frac{\partial J(\hat{\boldsymbol{w}})}{\partial\hat{\boldsymbol{w}}}=\frac{1}{N}\sum_{i=1}^{N}\left(h(\hat{\boldsymbol{x}}^{(i)})-y^{(i)}\right)\hat{\boldsymbol{x}}^{(i)}=\frac{1}{3}\left\{\left(\frac{1}{1+e^{-[1,0,-0.5][1,0,1]^T}}-1\right)\begin{bmatrix}1\\0\\1\end{bmatrix}+\left(\frac{1}{1+e^{-[1,0,-0.5][0,1,1]^T}}-1\right)\begin{bmatrix}0\\1\\1\end{bmatrix}+\left(\frac{1}{1+e^{-[1,0,-0.5][1,1,1]^T}}-0\right)\begin{bmatrix}1\\1\\1\end{bmatrix}\right\}$$

$$=[0.082,0,-0.13]^T$$

更新模型参数：

$$\hat{\boldsymbol{w}}^{(1)}=\hat{\boldsymbol{w}}^{(0)}-\rho\frac{\partial J(\hat{\boldsymbol{w}})}{\partial\hat{\boldsymbol{w}}}=\begin{bmatrix}1\\0\\-0.5\end{bmatrix}-0.2\times\begin{bmatrix}0.082\\0\\-0.13\end{bmatrix}=\begin{bmatrix}0.98\\0\\-0.47\end{bmatrix}$$

迭代一次后的决策面方程为：

$$0.98x_1-0.47=0$$

$$x_1\approx0.48$$

CHAPTER FOUR

第四章 支持向量机

Section 4, SVM

第四章 支持向量机

▶ 填空题

- 50个样本训练集上的线性SVM原始优化问题，具有（50）个（线性不等式）约束条件.
  - 解析：$y_i(\omega^T x_i + \gamma) - 1 \geq 0, i = 1, \ldots, 50$
- 线性SVM问题经过拉格朗日对偶处理后的优化问题的自变量是（拉格朗日乘子α_i）
  - 解析：$\max_{\alpha_i, i=1,\ldots,m} \left[\sum_{i=1}^m \alpha_i - \frac{1}{2}\sum_{i=1}^m \sum_{j=1}^m \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j\right]$

▶ 判断题

- 所有支撑向量均在间隔区域的边界上。 ×
  - 解析：没有声明SVM算法类型，软间隔SVM中支撑向量可以在间隔中
- 线性SVM求解必须依靠拉格朗日对偶技巧。 ×
  - 解析：拉格朗日对偶可以简化线性SVM，但原理上直接搜索也可以求解。

第四章 支持向量机

▶ 选择题

- 对于包含100个样本的训练集，线性SVM原始问题的待优化变量数量为（ ）

A. 2个　　B. 3个　　C. 101个　　D．无法确定

  - 解析：样本的特征维度未知

- 利用标准SMO算法求解SVM问题时，每轮迭代选取（ ）个系数变量进行更新

A. 1个　　B. 2个　　C. 样本数量　　D. 无法确定

  - 解析：每轮从所有样本数量个α_i中选取两个破坏KKT条件最严重的系数更新

▶ 简答题

- 在一个松弛项系数$C=0.3$的软间隔SVM模型中，如果一个样本$\boldsymbol{x}_i$对应的系数$\alpha_i=0.1$，试分析该样本与最优决策面和间隔区域的关系。

答：

- 样本$\boldsymbol{x}_i$对应的系数$\alpha_i=0.1$，即有$\alpha_i>0$,

- 根据KKT条件$\alpha_i(1-\xi_i-y_i f(x_i))=0$，推出$1-\xi_i-y_i f(x_i)=0$，因此$\boldsymbol{x}_i$是一个支撑向量。

- 其次，由于$0.1=\alpha_i<C=0.3$，根据KKT条件，$C=\alpha_i+\beta_i$推出$\beta_i>0$,

- 再次，根据KKT条件$\beta_i\xi_i=0$，推出$\xi_i=0$，即松弛变量为0。

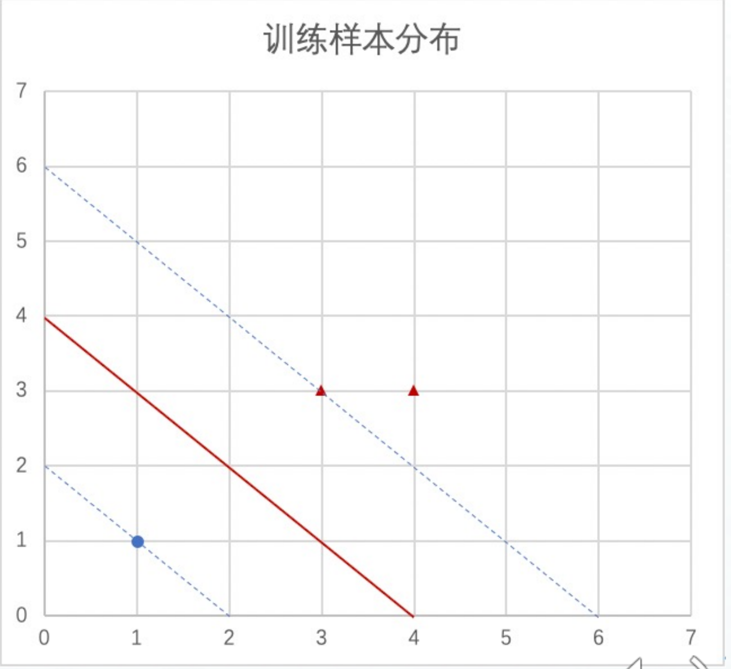
结论：样本x_i正好处在间隔区域的边界上。

计算题

- 已知一个如下图所示的训练数据集，其正类样本为$\boldsymbol{x_1} = (\boldsymbol{3,3})^{\boldsymbol{T}}$,$\boldsymbol{x_2} = (\boldsymbol{4,3})^{\boldsymbol{T}}$，负类样本是$\boldsymbol{x_3} = (\boldsymbol{1,1})^{\boldsymbol{T}}$。
- (1) 试写出上述问题的线性SVM原始优化问题的数学形式，包含目标函数与约束条件；
- 答：

$$\min \frac{1}{2}\|\boldsymbol{\omega}\|^2$$

$$\begin{aligned} \text{s.t.} \quad & 3\omega_1 + 3\omega_2 + \gamma \geq 1 \\ & 4\omega_1 + 3\omega_2 + \gamma \geq 1 \\ & \omega_1 + \omega_2 + \gamma \leq -1 \end{aligned}$$



▶ 计算题

- (2) 试写出经过拉格朗日对偶处理后的优化问题的数学形式，包含目标函数与约束条件；

答：

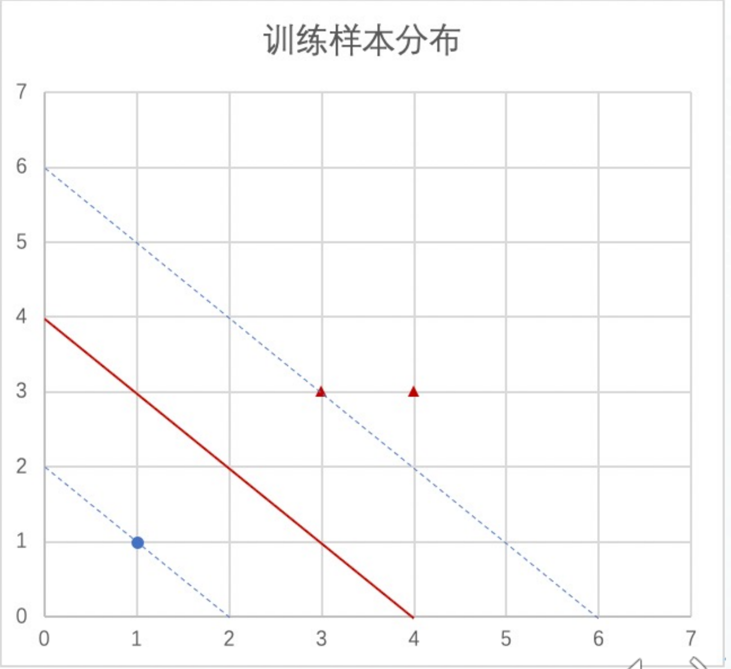
$$\max \sum_{i=1}^{3} \alpha_i - \frac{1}{2}\sum_{i=1}^{3}\sum_{j=1}^{3} \alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$

$$s.t., \quad \sum_{i=1}^{3} \alpha_i y_i = 0; \ \ \alpha_i \geq 0, \quad i = 1,2,3$$

代入$y_1 = 1, y_2 = 1, y_3 = -1; \boldsymbol{x}_1 = [3,3]^T, \boldsymbol{x}_2 = [4,3]^T, \boldsymbol{x}_3 = [1,1]^T$

$$\max(\alpha_1 + \alpha_2 + \alpha_3 - 9\alpha_1^2 - 12.5\alpha_2^2 - \alpha_3^2 - 21\alpha_1\alpha_2 + 7\alpha_2\alpha_3 + 6\alpha_3\alpha_1)$$

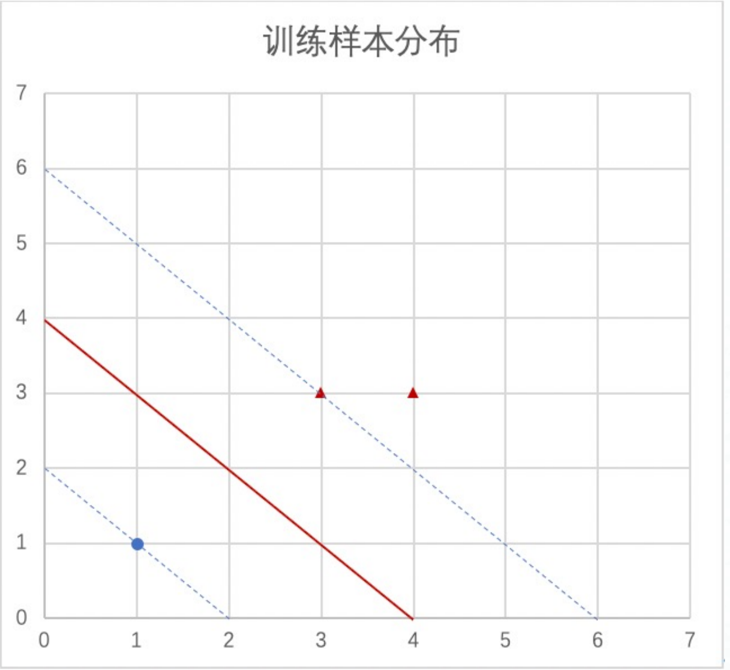
$$s.t., \quad \alpha_1 + \alpha_2 - \alpha_3 = 0; \ \ \alpha_i \geq 0, \quad i = 1,2,3$$



第四章 支持向量机

▶ 计算题

- (2)试写出经过拉格朗日对偶处理后的3条KKT条件。



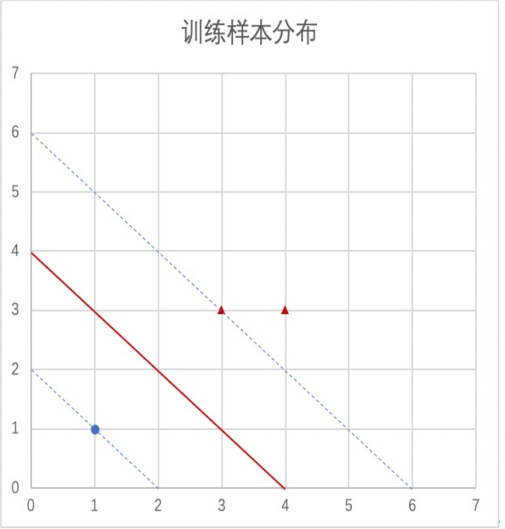
答：对于每个样本x_i,应满足以下三条KKT条件

$$\begin{cases} \alpha_i \geq 0 \\ y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i+\gamma)-1 \geq 0 \\ \alpha_i(y_i(\boldsymbol{\omega}^T\boldsymbol{x}_i+\gamma)-1)=0 \end{cases}, \quad i=1,2,3$$

代入$y_1=1, y_2=1, y_3=-1; \boldsymbol{x}_1=[3,3]^T, \boldsymbol{x}_2=[4,3]^T, \boldsymbol{x}_3=[1,1]^T$

得到三组KKT条件如下：

$$\begin{cases} \alpha_1 \geq 0 \\ 3\omega_1+3\omega_2+\gamma-1 \geq 0 \\ \alpha_1(3\omega_1+3\omega_2+\gamma-1)=0 \end{cases}, \begin{cases} \alpha_2 \geq 0 \\ 4\omega_1+3\omega_2+\gamma-1 \geq 0 \\ \alpha_2(4\omega_1+3\omega_2+\gamma-1)=0 \end{cases}, \begin{cases} \alpha_3 \geq 0 \\ -\omega_1-\omega_2-\gamma-1 \geq 0 \\ \alpha_3(\omega_1+\omega_2+\gamma+1)=0 \end{cases}$$



计算题

- (4) 对照图中的决策面和间隔区域，试推算样本$\boldsymbol{x}_i, i=1,2,3$对应的系数$\alpha_i, i=1,2,3$。

答：首先样本x_2位于间隔之外，所以有$4\omega_1+3\omega_2+\gamma-1>0$

根据（3）中的第二组KKT条件$\alpha_2(4\omega_1+3\omega_2+\gamma-1)=0$，可得$\alpha_2=0$；将$\alpha_2=0$代入（2）中的对偶问题中，消掉所有含有$\alpha_2$的项，得到：

$$\max(\alpha_1+\alpha_3-9\alpha_1^2-\alpha_3^2+6\alpha_3\alpha_1), \quad s.t. \quad \alpha_1-\alpha_3=0; \ \alpha_i\geq 0, i=1,3$$

由于x_1, x_3在间隔边界上，是支撑向量，所以有$\alpha_1=\alpha_3>0$

根据拉格朗日对偶过程中，广义拉格朗日函数对α_i求导为0，推出：

$$\boldsymbol{\omega}=\sum_{i=1}^{m}\alpha_i y_i \boldsymbol{x}_i$$

代入$\alpha_1=\alpha_3>0, \alpha_2=0$, 有：

$$\boldsymbol{\omega}=\alpha_1*1*\begin{bmatrix}3\\3\end{bmatrix}+\alpha_1*(-1)*\begin{bmatrix}1\\1\end{bmatrix}=\begin{bmatrix}2\alpha_1\\2\alpha_1\end{bmatrix}, \quad \omega_1=\omega_2=2\alpha_1$$

代入样本x_1, x_2对应的约束条件(支撑向量，等号成立）：$3\omega_1+3\omega_2+\gamma=1$；$\omega_1+\omega_2+\gamma=-1$

解得$\alpha_1=\alpha_3=\frac{1}{4}, \alpha_2=0$

CHAPTER FIVE

第五章 决策树

Section 5, Decision Tree

第五章 决策树

▶ 填空题

- 大多数决策树的节点类型可以分为（内部节点）和（叶子节点）。大多数根节点在类型上属于（内部）节点，剪枝操作剪去的是（叶子）节点

解析：具有子节点的根节点是否属于内部节点是一个定义问题，并不重要；但从面向对象编程的角度上看，将根节点作为内部节点处理是有意义的。

▶ 判断题

- 基于基尼指数的分裂结果与基于信息增益比的分裂结果总是相同的。×
  - 解析：基于基尼指数与基于信息增益选择的最优分裂特征在大多数情况下是相同的，但在理论上并不完全相同，有较小的可能会选出不同的分裂特征。基尼指数与信息增益比之间的差别就更大了。

第五章 决策树

选择题

- 假设离散特征A所有可能的取值为$\{a_1, a_2, a_3\}$, 设当前节点样本集为S, 且有$A(\boldsymbol{x}_i) \neq a_2, \forall \boldsymbol{x}_i \in S$，则利用特征A进行多叉树分裂后产生几个子节点？

  A. 2个　　B. 3个　　C. 无法确定　　D．无法分裂

  - 解析：虽然节点内训练样本的A特征取值都不等于a_2，但在多叉树分裂时仍需要保留$A(\boldsymbol{x}_i) = a_2$的子节点，这样才能在未来的测试中给出全面的分类结果。

- 当前节点内样本集包含5个样本，特征数量为3，其中两个离散特征，1个连续特征，采用多叉树方案，则当前节点共有多少种可选的分裂方案？

  A. 1种　　B. 2种　　C. 3种　　D. 8种

  - 解析：如采用多叉树策略，无论连续还是离散特征，每个特征对应一种分裂方案。

第五章 决策树

▶ 简答题

- 什么情况下一个叶子节点中会没有样本，此时该叶子节点返回的类别标签如何确定。

答：

当该叶子节点的父节点对应的样本集中没有符合该叶子节点特征取值要求的样本，但分裂特征的所有可能取值的集合中又包含了该叶子节点对应的特征取值时，该叶子节点中没有样本。

此时，该叶子节点返回其父节点中样本占比最多的类别标签。

第五章 决策树

计算题

- Iris数据库的某个特征增广版本包含7个样本，具体情况如表格所示。对比香气和颜色两种特征，分别依据信息增益、信息增益比和基尼指数给出分裂特征选择结果及计算过程。

序号	香气	颜色	花萼长度	花萼宽度	花瓣长度	花瓣宽度	类别
1	有	红	5.1	3.5	1.4	0.2	setosa
2	有	红	4.9	3	1.4	0.2	setosa
3	有	粉	4.7	3.2	1.3	0.2	setosa
4	有	紫	5.3	3.7	1.5	0.2	setosa
5	无	粉	7	3.2	4.7	1.4	versicolor
6	无	紫	6.4	3.2	4.5	1.5	versicolor
7	无	紫	6.3	3.3	6	2.5	virginica
8	有	紫	5.8	2.7	5.1	1.9	virginica

(1) 根据信息增益选择分裂特征

\- 根节点的经验熵：$\hat{H}(Y)=-\sum_{j=1}^{3}P(y=\omega_j)\log P(y=\omega_j)=-\left(\frac{4}{8}\log_2\frac{4}{8}+\frac{2}{8}\log_2\frac{2}{8}+\frac{2}{8}\log_2\frac{2}{8}\right)=1.5$

\- 香气属性

根据香气特征划分子集集合：$S_1=\{1(\mathrm{se}),2(\mathrm{se}),3(\mathrm{se}),4(\mathrm{se}),8(\mathrm{vi})\}$, $S_2=\{5(\mathrm{ve}),6(\mathrm{ve}),7(\mathrm{vi})\}$

计算每个子节点的信息熵：$\hat{H}(S_1)=-\left(\frac{4}{5}\log_2\frac{4}{5}+\frac{1}{5}\log_2\frac{1}{5}\right)=0.722, \hat{H}(S_2)=-\left(\frac{2}{3}\log_2\frac{2}{3}+\frac{1}{3}\log_2\frac{1}{3}\right)=0.918$

加权平均经验熵：$\hat{H}_1=\sum_{k=1}^{2}\left(\frac{N_k}{N}\right)\hat{H}(S_k)=\frac{5}{8}\times0.722+\frac{3}{8}\times0.918=0.796$

属性“香气”信息增益为：$G_1=\hat{H}(Y)-\hat{H}_1=1.5-0.796=0.704$

\- “颜色”属性

根据颜色属性划分子集集合：$S_1=\{1(\mathrm{se}),2(\mathrm{se})\}$, $S_2=\{3(\mathrm{se}),5(\mathrm{ve})\}$, $S_3=\{4(\mathrm{se}),6(\mathrm{ve}),7(\mathrm{vi}),8(\mathrm{vi})\}$

计算每个子节点的信息熵：$\hat{H}(S_1)=0, \hat{H}(S_2)=-\left(\frac{1}{2}\log_2\frac{1}{2}+\frac{1}{2}\log_2\frac{1}{2}\right)=1.0$， $\hat{H}(S_3)=-\left(\frac{1}{4}\log_2\frac{1}{4}+\frac{1}{4}\log_2\frac{1}{4}+\frac{2}{4}\log_2\frac{2}{4}\right)=1.5$

加权平均经验熵：$\hat{H}_2=\sum_{k=1}^{3}\left(\frac{N_k}{N}\right)\hat{H}(S_k)=\frac{2}{8}\times0+\frac{2}{8}\times1.0+\frac{4}{8}\times1.5=1.0$

信息增益：$G_2=\hat{H}(Y)-\hat{H}_2=1.5-1.0=0.5$

结论1：根据信息增益$G_1>G_2$，所以选择“香气”属性分裂特征。

计算题

- Iris数据库的某个特征增广版本包含7个样本，具体情况如表格所示。对比香气和颜色两种特征，分别依据信息增益、信息增益比和基尼指数给出分裂特征选择结果及计算过程。

序号	香气	颜色	花萼长度	花萼宽度	花瓣长度	花瓣宽度	类别
1	有	红	5.1	3.5	1.4	0.2	setosa
2	有	红	4.9	3	1.4	0.2	setosa
3	有	粉	4.7	3.2	1.3	0.2	setosa
4	有	紫	5.3	3.7	1.5	0.2	setosa
5	无	粉	7	3.2	4.7	1.4	versicolor
6	无	紫	6.4	3.2	4.5	1.5	versicolor
7	无	紫	6.3	3.3	6	2.5	virginica
8	有	紫	5.8	2.7	5.1	1.9	virginica

(2) 根据信息增益比选择分裂特征

“香气”属性经验熵：

$$\hat{H}(\text{"香气"}) = -\sum_{k=1}^{2}\frac{N_k}{N}\log\frac{N_k}{N} = -\left(\frac{5}{8}\log\frac{5}{8}+\frac{3}{8}\log\frac{3}{8}\right) = 0.954$$

信息增益比为：

$$R_1 = \frac{0.704}{0.954} = 0.738$$

“颜色”属性经验熵：

$$\hat{H}(\text{"颜色"}) = -\sum_{k=1}^{3}\frac{N_k}{N}\log\frac{N_k}{N} = -\left(\frac{2}{8}\log\frac{2}{8}+\frac{2}{8}\log\frac{2}{8}+\frac{4}{8}\log\frac{4}{8}\right) = 1.5$$

信息增益比为：

$$R_2 = \frac{0.5}{1.5} = 0.33$$

信息增益比$R_1 > R_2$，所以选择“香气”属性进行特征分裂

计算题

- Iris数据库的某个特征增广版本包含7个样本，具体情况如表格所示。对比香气和颜色两种特征，分别依据信息增益、信息增益比和基尼指数给出分裂特征选择结果及计算过程。

序号	香气	颜色	花萼长度	花萼宽度	花瓣长度	花瓣宽度	类别
1	有	红	5.1	3.5	1.4	0.2	setosa
2	有	红	4.9	3	1.4	0.2	setosa
3	有	粉	4.7	3.2	1.3	0.2	setosa
4	有	紫	5.3	3.7	1.5	0.2	setosa
5	无	粉	7	3.2	4.7	1.4	versicolor
6	无	紫	6.4	3.2	4.5	1.5	versicolor
7	无	紫	6.3	3.3	6	2.5	virginica
8	有	紫	5.8	2.7	5.1	1.9	virginica

(3) 根据基尼指数选择分裂特征

计算“香气”属性的信息增益：

$$Gini(\text{“香气 = 有”}) = 1 - \left(\frac{4}{5}\right)^2 - \left(\frac{1}{5}\right)^2 = 0.320\text{，}Gini(\text{"香气 = 无"}) = 1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2 = 0.444$$

$$Gini\_score(\text{"香气"}) = \frac{5}{8}\times 0.320 + \frac{3}{8}\times 0.444 = 0.37$$

计算“颜色”属性的信息增益：

$$Gini(\text{“颜色 = 红”}) = 1 - \left(\frac{2}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0\text{，}Gini(\text{“颜色 = 粉”}) = 1 - \left(\frac{1}{2}\right)^2 - \left(\frac{1}{2}\right)^2 = 0.5\text{，}Gini(\text{"颜色 = 紫"}) = 1 - \left(\frac{1}{4}\right)^2 - \left(\frac{1}{4}\right)^2 - \left(\frac{2}{4}\right)^2 = 0.625$$

$$Gini\_score(\text{"颜色"}) = \frac{2}{8}\times 0 + \frac{2}{8}\times 0.5 + \frac{4}{8}\times 0.625 = 0.437$$

由于$Gini_score(\text{"香气"}) < Gini_score(\text{"颜色"})$，所以按照“香气”特征进行分裂。

CHAPTER SIX

第六章 集成学习

Section 6, Ensemble Learning

▶ 填空题

- 如果将多项式看作一个集成学习系统，其个体学习器是（幂/多项式）函数，个体学习器的参数是（多项式系数及常数项）。

  解析：个体学习器可以是单项式（幂函数）也可以是多项式，只要组合在一起能够表达为多项式即可。

▶ 判断题

- Adaboost算法中，被个体学习器h_1错误分类的样本的权重在h_2的学习中一定会增加。 √
  - 解析：样本x_i在h_1中的权重为$\omega_{1i} = \frac{1}{N}$, 由于学习器为弱分类算法，误差$\epsilon_1 < 0.5$,因此$\alpha_1 > 0$, 对于错分样本$x_i$, $\omega_{2,i} = \frac{\omega_{1i}}{Z_1} e^{-\alpha_1 f(x) h_1(x)}$，显然错分样本的相对权重增加了，因此样本权重在归一化后必然大于平均值$\frac{1}{N}$。

第六章 集成学习

▶ 选择题

- 设样本集为$X = \{1,3,5,7\}, Y = \{+1,-1,-1,+1\}$，权重为$\mathcal{D}_t = \{0.1,0.5,0.2,0.2\}$，个体分类器函数形式为$h_t(x) = \text{sign}(x - v)$。则当前$v$的最优取值为：

A. 0　　B. 2　　C. 4　　D . 6

  - 解析：当$v = 6$时，只有第一个样本$x = 1, y = +1$被错分，加权误差$\epsilon = 0.1$

▶ 简答题

- 简述集成学习中个体学习器的设计与学习要满足什么样的规则，并加以解释。
- 答：集成学习系统中的个体学习器设计与学习应满足“好而不同”的原则。
- “好”是指每个个体分类器的误差应小于0.5，即满足弱学习器的基本要求；
- “不同”是指个体学习器的识别结果应存在多样性，以提升集成学习的泛化性能。

序号	1	2	3	4	5	6
x	0	1	2	3	4	5
y	1	1	-1	-1	1	1

计算题

- 给定如表所示训练数据集。假设弱分类器由$x<v$或$x>v$产生，试使用AdaBoost算法求解。

（1）初始化样本权重：$\omega_i = \frac{1}{N} = \frac{1}{6}$

故上述训练集D所对应的权重集合：$D_1 = \{\omega_{1,1}, \omega_{1,2}, \omega_{1,3}, \omega_{1,4}, \omega_{1,5}, \omega_{1,6}\} = \{\frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}, \frac{1}{6}\}$

由题可知，以序号为标准，阈值v可取-0.5，0.5，1.5，2.5，3.5，4.5，5.5。

分类误差率为样本 $\boldsymbol{x}$ 的预测分类 $\mathrm{h}_t(\boldsymbol{x})$ 和实际分类 $f(\boldsymbol{x})$ 不相同的概率：$\epsilon_1 = P_{x\sim D_t}(\mathrm{h}_t(\boldsymbol{x}) \neq f(\boldsymbol{x}))$

故选择分类器$x > v$，阈值 $v = -0.5$，对应的分类误差 $e = 2/6 < 0.5$

此时的分类器权重：$\alpha_1 = \frac{1}{2}\ln\frac{1-\epsilon_1}{\epsilon_1} = \frac{1}{2}\ln\frac{1-\frac{1}{3}}{\frac{1}{3}} = \ln\sqrt{2}$

规范化因子Z_i的计算如下，它使各个值的新权重值之和为1:

$$Z_1 = \sum_{i=1}^{6} \omega_{1,i} e^{-\alpha_1 f(x) h_1(x)} = \frac{1}{6}\mathrm{e}^{-\ln\sqrt{2}*1*(-1)}\times 2 + \frac{1}{6}\mathrm{e}^{-\ln\sqrt{2}*1*1}\times 4 = \frac{2\sqrt{2}}{3}$$

被错误分类的样本的权重($i = 3{,}4$)：$\omega_{2,i} = \frac{\omega_{1i}}{Z_1} e^{-\alpha_1 f(x) h_1(x)} = \frac{\frac{1}{6}}{\frac{2\sqrt{2}}{3}} e^{-\ln\sqrt{2}*1*(-1)} = \frac{1}{4}$

被正确分类的样本的权重($i = 1{,}2{,}5{,}6$)：$\omega_{2,i} = \frac{\omega_{1i}}{Z_1} e^{-\alpha_1 f(x) h_1(x)} = \frac{\frac{1}{6}}{\frac{2\sqrt{2}}{3}} e^{-\ln\sqrt{2}*1*1} = \frac{1}{8}$

此时更新的权重集合：$D_2 = \{\omega_{2,1}, \omega_{2,2}, \omega_{2,3}, \omega_{2,4}, \omega_{2,5}\} = \{\frac{1}{8}, \frac{1}{8}, \frac{1}{4}, \frac{1}{4}, \frac{1}{8}, \frac{1}{8}\}$

v	$x < v$错分样本	$x > v$错分样本	最小分类误差e
-0.5	1,2,5,6	3,4	2/6
0.5	2,5,6	1,3,4	3/6
1.5	5,6	1,2,3,4	2/6
2.5	3,5,6	1,2,4	3/6
3.5	3,4,5,6	1,2	2/6
4.5	3,4,6	1,2,5	3/6
5.5	3,4	1,2,5,6	2/6

序号	1	2	3	4	5	6
x	0	1	2	3	4	5
y	1	1	-1	-1	1	1

计算题

- 给定如表所示训练数据集。假设弱分类器由$x<v$或$x>v$产生，试使用AdaBoost算法求解。

（2）权重更新之后，错分样本和加权误差情况如右表

根据加权分类误差，选择分类器$\boldsymbol{x}<\boldsymbol{v}$，阈值$v=1$，$\alpha_2=\frac{1}{2}ln\frac{1-\epsilon_1}{\epsilon_1}=\frac{1}{2}ln\frac{1-1/4}{1/4}=ln\sqrt{3}$

归一化因子：$Z_2=\sum_{j=1}^{6}\omega_{2,j}e^{-\alpha_2 f(x)G_2(x_i)}=2*\frac{1}{8}*\sqrt{3}+2*\frac{1}{4}*\frac{\sqrt{3}}{3}+2*\frac{1}{8}*\frac{\sqrt{3}}{3}=\frac{\sqrt{3}}{2}$

被错误分类的样本的权重($i=5,6$)：$\omega_{3,i}=\frac{\omega_{2,i}}{Z_1}e^{-\alpha_2 f(x)h_2(x)}=\frac{\frac{\sqrt{3}}{8}}{\frac{\sqrt{3}}{2}}=\frac{1}{4}$

上次被错误分类，此次被正确分类的样本的权重($i=3,4$)：$\omega_{3,4}=\frac{\omega_{2,1}}{Z_1}e^{-\alpha_1 f(x)h_2(x)}=\frac{\frac{\sqrt{3}}{12}}{\frac{\sqrt{3}}{2}}=\frac{1}{6}$

上次被正确分类，此次仍被正确分类的样本的权重($i=1,2$)：$\omega_{3,i}=\frac{\omega_{2,i}}{Z_1}e^{-\alpha_i f(x)h_2(x)}=\frac{\frac{\sqrt{3}}{24}}{\frac{\sqrt{3}}{2}}=\frac{1}{12}$

此时更新的权重集合：$D_3=\{\omega_{3,1},\omega_{3,2},\omega_{3,3},\omega_{3,4},\omega_{3,5},\omega_{3,6}\}=\{\frac{1}{12},\frac{1}{12},\frac{1}{6},\frac{1}{6},\frac{1}{4},\frac{1}{4}\}$

故两轮训练后，弱分类器输出：$H(\boldsymbol{x})=sign(ln\sqrt{2}*\mathrm{sign}(x+0.5)+ln\sqrt{3}*\mathrm{sign}(1.5-\mathrm{x}))$

v	$x<v$错分样本	$x>v$错分样本	最小分类误差e
-0.5	1,2,5,6	3,4	1/2
0.5	2,5,6	1,3,4	3/8
1.5	5,6	1,2,3,4	1/4
2.5	3,5,6	1,2,4	3/8
3.5	3,4,5,6	1,2	1/4
4.5	3,4,6	1,2,5	3/8
5.5	3,4	1,2,5,6	1/2

CHAPTER SEVEN

第七章 模型评估

Section 7, Model Evaluation

第七章 模型评估

▶ 填空题

- 当两个假设在训练集上具有完全相同的经验误差时，应依据（归纳偏好）来进行假设选择。
- 函数$h(x;a,b)=\mathrm{sign}(\mathrm{a}x-b), \mathrm{a},\mathrm{b}\in\Re$支撑的假设空间为$\mathcal{H}$, 则增长函数$\Pi_{\mathcal{H}}(3)=(6)$
  - 解析：等价于双向阈值函数
- 使用10-折交叉验证法处理Iris数据集，则每一次验证时训练集中的setosa类样本数量是（45）。验证集的样本数量是（15）

▶ 判断题

- 一个高效PAC可学问题一定是PAC可学问题。 √
- 模型在独立同分布训练集上的经验误差总是小于该模型在该分布上的泛化误差。 ×
  - 解析：虽然泛化误差上界大于经验误差，但真实的泛化误差不一定大于经验误差，当训练集加好都是一些难样本时，泛化误差可能会小于经验误差

第七章 模型评估

▶ 简答题

- 请简要阐述经验误差与泛化误差之间的关系。

  答： - 当训练样本数量N趋近于无穷大时，泛化误差等于经验误差；

  - 经验误差的数学期望等于泛化误差;

  - 利用经验误差与模型参数可以描述泛化误差的边界。

- 请结合本课程讲授的分类算法，说明三种具体的正则化方法。

  答： - 线性回归，在损失函数中加入正则化项$||\boldsymbol{w}||^2$,得到岭回归封闭解；

  - SVM，在线性SVM基础上加入松弛变量，构建软间隔SVM，形成结构风险项；

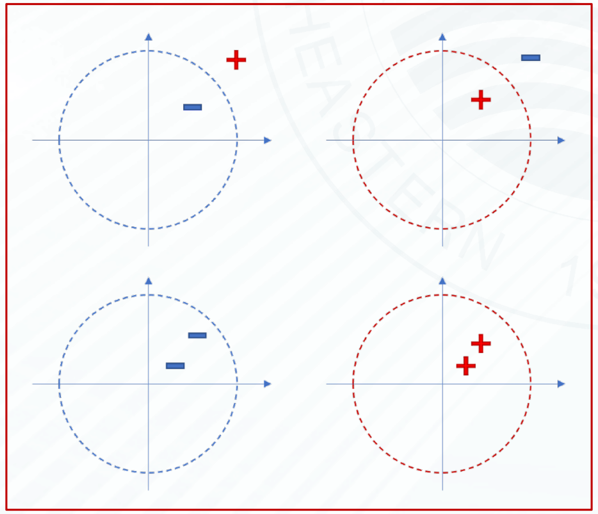
  - 决策树，加入剪枝操作，降低决策树的复杂度。

画图题

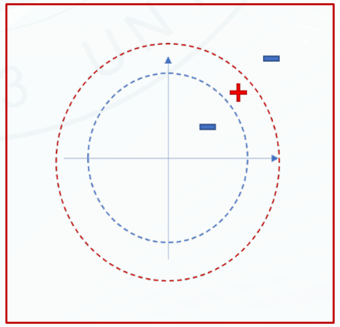
- 考虑一个二分类问题，$x \in \Re^2, y \in \{+1,-1\}$. 给定一个假设空间如下：$\mathcal{H} = \{h(x;a,b) = \text{sign}[a(x^T x - b)] | a \in \{+1,-1\}, b \in [0,+\infty)\}$。请通过画图法找出假设空间$\mathcal{H}$的VC维d的具体取值（提示：画出$\mathcal{H}$打散d个样本但无法打散d+1个样本的情况）

答：映射函数$\text{sign}[a(x^T x - b)]$的判别效果相当于在利用一个以圆点为中心，半径为$\sqrt{b}$的圆将样本空间$\mathcal{X}$分为两个区域，并根据a的数值决定区域的类别标签。当d=2时，有四种情况如下，（红圈表示圈内为正类，蓝圈表示圈内为负类）显然能被$\mathcal{H}$打散

当$d = 3$时，有四种情况如下，以下样本分布无法被任意$h \in \mathcal{H}$正确分类，因此，$d = 3$时样本集无法被假设空间$\mathcal{H}$打散



$d = 2$



$d = 3$

结论：假设空间$\mathcal{H}$的VC维$d = 2$。

样本序号	1	2	3	4	5	6	7	8	9	10
Y	P	N	N	P	P	P	N	P	N	P
H	P	N	P	P	N	P	P	P	N	N

计算题

- 已知测试集真实标签Y和分类器分类结果H如表所示，请计算以下性能指标，并给出计算方法。
- 真阳率、假阳率、查准率、查全率、F1度量

答：首先统计TP，FP，TN，FN的样本数量，如右表

既 $TP=4, FP=2, TN=2, FN=2$

	T	F
P	4	2
N	2	2

相应指标计算如下：

真阳率：$TPR=\frac{TP}{TP+FP}=\frac{4}{4+2}\approx 0.67$

假阳率：$FPR=\frac{FP}{FP+TN}=\frac{2}{2+2}\approx 0.5$

查准率：Precision$=\frac{TP}{TP+FP}=\frac{4}{4+2}\approx 0.67$

查全率：Recall$=TPR\approx 0.67$

F1度量：$F_1=\frac{2*Precision*Recall}{Precision+Recall}=\frac{2*\frac{2}{3}*\frac{2}{3}}{\frac{2}{3}+\frac{2}{3}}=\frac{2}{3}\approx 0.67$

CHAPTER EIGHT

第八章 特征选择与学习

Section 8, Feature Selection and Learning

第八章 特征选择与学习

▶ 填空题

- MDS和IsoMap的相同之处是（降维前后样本间距离保持不变），不同之处是前者使用（欧式）距离，后者使用（测地线）距离
- 10维空间中的5个样本，最多需要（4）维的子空间可以保证投影损失为0。
  - 解析：三维空间中的2个样本，可以嵌入到一条直线上（1维）。

▶ 判断题

- IsoMap流形学习算法无法直接实现训练集以外的新样本的特征提取。 √
  - 解析：IsoMap的测地线距离和距离矩阵都是针对训练数据的。
- 稀疏编码算法的字典中的特征向量必须正交。 ×
  - 解析：一般不能正交，因为满足正交性的基向量很少，无法满足编码的稀疏性

简答题

- 请简要列举3种可以用于身份识别的生物特征，并分析其特点
- 答：　- 指纹特征：准确性高、识别成本低、采集成本中等

  - 人脸特征：准确性较高、识别成本中等、采集成本低

  - 基因特征：准确性极高、识别成本高、采集成本高

- 给出PCA中方差最大化和投影损失最大化两种思路在第一个基向量上求取最优解的等价性证明。

答：投影方差的目标函数：$J_1(\boldsymbol{b}_1) = E\{y_1^2\} = E\{(\boldsymbol{b}_1^T\boldsymbol{x})^2\} = \boldsymbol{b}_1^T E\{\boldsymbol{x}\boldsymbol{x}^T\}\boldsymbol{b}_1 \approx \boldsymbol{b}_1^T C_X \boldsymbol{b}_1$

投影损失的目标函数：$J_2(\boldsymbol{b}_1) = E\{\|\boldsymbol{x} - y_1\boldsymbol{b}_1\|^2\} = E\{(\boldsymbol{x} - (\boldsymbol{b}_1^T\boldsymbol{x})\boldsymbol{b}_1)^T(\boldsymbol{x} - (\boldsymbol{b}_1^T\boldsymbol{x})\boldsymbol{b}_1)\} = E\{\boldsymbol{x}^T\boldsymbol{x}\} - E\{2(\boldsymbol{b}_1^T\boldsymbol{x})^T(\boldsymbol{b}_1^T\boldsymbol{x})\} + E\{\boldsymbol{b}_1^T(\boldsymbol{b}_1^T\boldsymbol{x})^T(\boldsymbol{b}_1^T\boldsymbol{x})\boldsymbol{b}_1\} = E\{\boldsymbol{x}^T\boldsymbol{x}\} - E\{(\boldsymbol{b}_1^T\boldsymbol{x})^2\} \approx C_X - \boldsymbol{b}_1^T C_X \boldsymbol{b}_1$

由于C_X是一个常数，因此：$\boldsymbol{b}_1^* = \underset{\boldsymbol{b}_1}{\operatorname{argmax}} J_1(\boldsymbol{b}_1) = \underset{\boldsymbol{b}_1}{\operatorname{argmin}} J_2(\boldsymbol{b}_1)\ s.t.\|\boldsymbol{b}_1\| = 1$

说明两者在具有完全相同的第一个基向量$\boldsymbol{b}_1$的最优解。

CHAPTER NINE

第九章 神经网络

Section 9, Neural Networks

第九章 神经网络

▶ 填空题

• 单层感知器无法解决的代表性问题是（ XOR问题 ）。

▶ 简答题

• 请从集成学习的角度解释MLP为何能解决XOR问题。

答：典型的MLP的每一个隐层神经元相当于一个单层感知器算法，可以作为个体分类器；隐层向输出层的连接与信号计算相当于STACK集成学习中的次级学习器。

虽然单一的SLP无法解决XOR问题，但多个SLP可以在样本空间中画出任意形状的区域，使得区域内和区域外分属于两类，因此能够解决XOR问题。

▶计算题

- 对于一个5输入的MP神经元模型，采用Logistic Sigmoidal激活函数，输入样本为 $[0.2, 0.5, -0.3, -0.7, 0.6]$，对应连接权值为 $[0.2, 0.3, -0.1, -0.5, 0.9]$，偏置为-1，请计算该MP神经元模型的输出，并给出计算过程。

解：设权重 $w = [0.2, 0.3, -0.1, -0.5, 0.9]^T$，样本 $\boldsymbol{x} = [0.2, 0.5, -0.3, -0.7, 0.6]^T$，偏置为 $b = -1$；

则有：$y = \boldsymbol{w}^T\boldsymbol{x} + b = [0.2, 0.3, -0.1, -0.5, 0.9] \times \begin{bmatrix} 0.2 \\ 0.5 \\ -0.3 \\ -0.7 \\ 0.6 \end{bmatrix} - 1 = 0.11$

MP神经元输出为：$S(y) = \frac{1}{1+\exp(-y)} = \frac{1}{1+\exp(-0.11)} \approx 0.527$