清華大學
i-VisionGroup

《模式识别与机器学习》第1次习题课

助教：许修为、黄原辉

2023年11月12日

第二章习题

填空题

- 注意：概念题用尽量精简的话进行回答，可以参考教材的定义；简答题意思正确即可

一、填空题

1. LDA 线性判别分析的求解过程是将 N 维特征向量投影在________维空间中进行。

答案：不超过 C-1，C 为类别总数

2. 对数几率函数$\theta(x)=$ ____________，其导数$\theta'(x)=$ __________（可用$\theta(x)$表示）。3

答案：$\frac{1}{1+e^{-x}}$；$\theta(x)\left(1-\theta(x)\right)$

3. 散度越大，说明两类模式分布差异__________。

答案：越大

4. 散度是根据__________构造的可分性判据。

答案：类概率密度

5. 线性判别函数的正负和数值大小的几何意义是______________________。

答案：正（负）表示样本点位于判别界面法向量指向的正（负）半空间中；绝对值正比于样本点到判别界面的距离

第二章习题

□ 计算题

- **注意：**题目较为简单，直接代入公式即可。最后结果尽量用分数表示，或者用高精度（3-4位）小数，避免因为精度太低丢分

二、计算题

两类数据如下所示：$X_1=(x_1,x_2)=\{(4,1),(2,4),(2,3),(3,6),(4,4)\}$，$X_2=(x_1,x_2)=\{(9,10),(6,8),(9,5),(8,7),(10,8)\}$，计算 LDA 的投影方向。

答案：

计算类均值：

$$\mu_1=[3.0,\ 3.6]^T\quad \mu_2=[8.4,\ 7.6]^T$$

计算类内散度矩阵：

$$S_w=\Sigma_0+\Sigma_1=\sum_{x\in X_1}(\boldsymbol{x}-\boldsymbol{\mu}_1)(\boldsymbol{x}-\boldsymbol{\mu}_1)^T+\sum_{x\in X_2}(\boldsymbol{x}-\boldsymbol{\mu}_2)(\boldsymbol{x}-\boldsymbol{\mu}_2)^T=\begin{bmatrix}13.2 & -2.2\\ -2.2 & 26.4\end{bmatrix}$$

计算 LDA 投影方向：

$$\boldsymbol{w}=\boldsymbol{S}_w^{-1}(\boldsymbol{\mu}_0-\boldsymbol{\mu}_1)=\left[-\frac{344}{781},\ -\frac{147}{781}\right]^T\approx[-0.44,\ -0.19]^T$$

第二章习题

证明题

三、证明题

设$x=[x_1,x_2]^T$为二维空间中的特征向量，二分类问题的线性判别函数为$g(x)=w^Tx+w_0$，其中权向量$w=[w_1,w_2]^T$是决策方程$g(x)=0$对应的超平面的法向量，其长度为$\|w\|$.

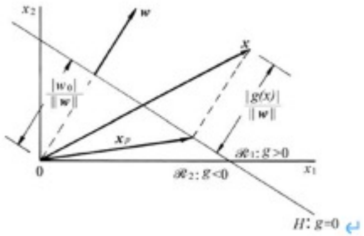
(1) 证明：从x到超平面$g(x)=0$的距离为：

$$r=\frac{|g(x)|}{\|w\|}$$

(2) 证明：x在超平面上的投影为：

$$x_p=x-\frac{|g(x)|}{\|w\|^2}w$$

答案：



(1) 设x到超平面的距离向量为x_d，x在超平面上的投影向量为x_p，则有

$$x=x_p+x_d$$

若x_d与w同向，则有

$$g(x)=w^Tx+w_0=w^T\left(x_p+r\frac{w}{\|w\|}\right)+w_0=g(x_p)+r\frac{w^Tw}{\|w\|}=r\|w\|$$

若x_d与w反向，则有

$$g(x)=w^Tx+w_0=w^T\left(x_p-r\frac{w}{\|w\|}\right)+w_0=g(x_p)-r\frac{w^Tw}{\|w\|}=-r\|w\|$$

(2) 设x到超平面的距离向量为x_d，x在超平面上的投影向量为x_p，则有

$$x=x_d+x_p$$

若x_d与w同向，则有

$$x_p=x-r\frac{w}{\|w\|}=x-\frac{g(x)}{\|w\|}\frac{w}{\|w\|}=x-\frac{|g(x)|}{\|w\|^2}w$$

若x_d与w反向，则有

$$x_p=x+r\frac{w}{\|w\|}=x+\frac{-g(x)}{\|w\|}\frac{w}{\|w\|}=x-\frac{|g(x)|}{\|w\|^2}w$$

第三章习题

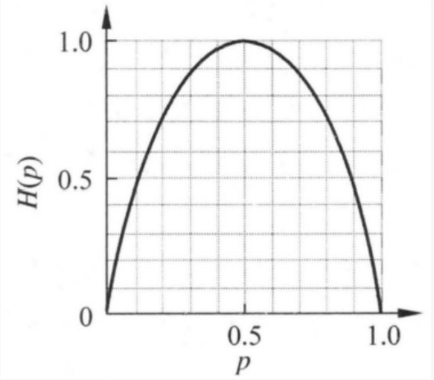
一、填空题（3分）

1. 构建决策树时，选择最优划分属性的准则有__________________。

答案：信息增益、增益率、基尼指数

2. 在决策树中，信息熵越大，证明样本集合的纯度越_____（高/低）。

答案：低　$H(X) = -\sum_{i=1}^{n} p_i \log_2 p_i$



3. 构建决策树的信息增益准则对取值数目更_________的属性有利。

答案：多

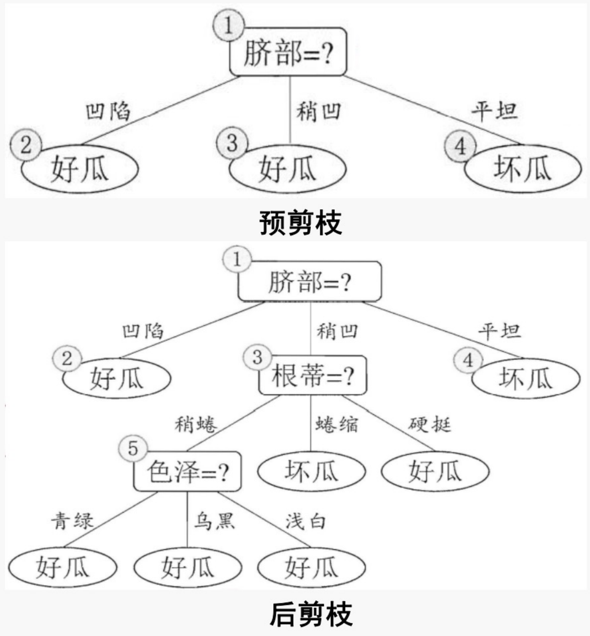
- 代表在一个条件下信息不确定性减少的程度
- 定义：属性 a 对样本集 D 的信息增益 $\mathrm{Gain}(D,a)$ 定义为样本集 D 的经验熵 $H(D)$ 与属性 a 给定条件下样本集 D 的经验条件熵 $H(D|a)$ 之差，即：

$$\mathrm{Gain}(D,a) = H(D) - H(D|a) = H(D) - \sum_{v=1}^{V} \frac{|D^v|}{|D|} H(D^v)$$

- 上式中，$\frac{|D^v|}{|D|}$ 表示属性 a 的边缘分布；$H(D^v)$ 表示属性 a 取值 a^v 时，样本集 D 的条件概率分布的熵

补充：

第三章习题



二、选择题（3分）

1. （单选题）以下关于决策树的说法错误的是（ B ）

   A. 决策树容易过拟合，可以使用预剪枝或后剪枝进行处理

   B. 相比于预剪枝，后剪枝更容易得到更简单的决策树

   C. 后剪枝需要判断根结点的剪枝与否

   D. 决策树既可用于分类问题，也可用于回归问题

2. （多选题）关于决策树方法，以下说法正确的是（ AD ）

   A. 多变量决策树能得到与坐标轴呈一定夹角的分类面

   B. 增益率准则对取值数目多的属性有所偏好

   C. 基尼指数衡量了数据在类别上的不确定性的减小程度

   D. 预剪枝和后剪枝策略都可以用于预防过拟合

$$\mathrm{GainRatio}(D,a)=\mathrm{Gain}(D,a)/\mathrm{IV}(a)$$

$$\mathrm{IV}(a)=-\sum_{v=1}^{V}\frac{|D^v|}{|D|}\log\frac{|D^v|}{|D|}$$

$$\mathrm{Gini}(D)=1-\sum_{k=1}^{|\mathcal{Y}|}p_k^2$$

$$\mathrm{GiniIndex}(D,a)=\sum_{v=1}^{V}\frac{|D^v|}{|D|}\mathrm{Gini}(D^v)$$

第三章习题

- 下表是由15个样本组成的贷款申请数据集，数据包括贷款申请人的年龄、收入情况、是否有车、信用情况四项属性，最后一列为是否同意贷款，作为我们的预测结果：

1. 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，并将{5，6，14，15}号样本作为验证集，其余样本作为训练集。请利用增益率和后剪枝策略建立决策树，写出决策树的建立过程并画出最终的决策树结构。
2. 将1中的增益率准则替换为基尼指数准则，后剪枝策略替换为预剪枝策略，请重新建立决策树，写出决策树的建立过程并画出最终的决策树结构。

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	59	否	良好	好	是
14	61	否	良好	非常好	是
15	29	是	良好	一般	是

第三章习题

解答：

1. 以划分根节点为例。首先计算各属性的增益率，如下：

$$Ent(D) = -\left(\frac{4}{11}log\frac{4}{11} + \frac{7}{11}log\frac{7}{11}\right) = 0.946$$

$$Gain(\text{年龄}) = Ent(D) - \left(-\frac{4}{11}\left(\frac{3}{4}log\frac{3}{4} + \frac{1}{4}log\frac{1}{4}\right) - \frac{4}{11} * 0 - \frac{3}{11}\left(\frac{1}{3}log\frac{1}{3} + \frac{2}{3}log\frac{2}{3}\right)\right) = 0.400$$

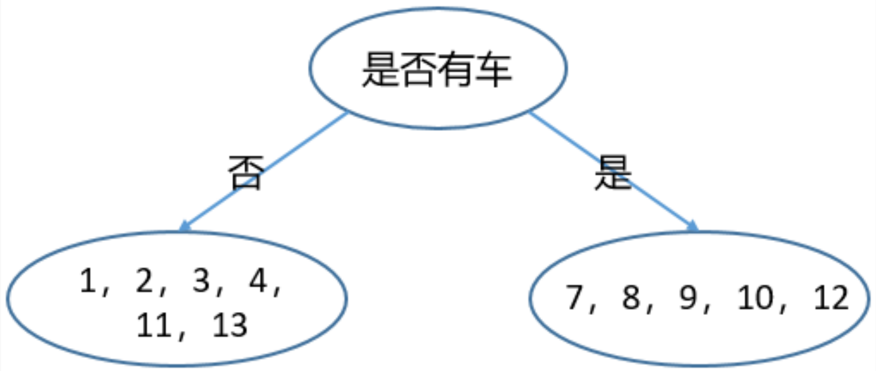
$$GainRatio(\text{年龄}) = \frac{Gain(\text{年龄})}{-\frac{4}{11}log\frac{4}{11} - \frac{4}{11}log\frac{4}{11} - \frac{3}{11}log\frac{3}{11}} = 0.254$$

$$GainRatio(\text{是否有车}) = \frac{Gain(\text{是否有车})}{-\frac{6}{11}log\frac{6}{11} - \frac{5}{11}log\frac{5}{11}} = 0.447$$

$$GainRatio(\text{收入情况}) = \frac{Gain(\text{收入情况})}{-\frac{8}{11}log\frac{8}{11} - \frac{3}{11}log\frac{3}{11}} = 0.258$$

$$GainRatio(\text{信用情况}) = \frac{Gain(\text{信用情况})}{-\frac{3}{11}log\frac{3}{11} - \frac{5}{11}log\frac{5}{11} - \frac{3}{11}log\frac{3}{11}} = 0.401$$

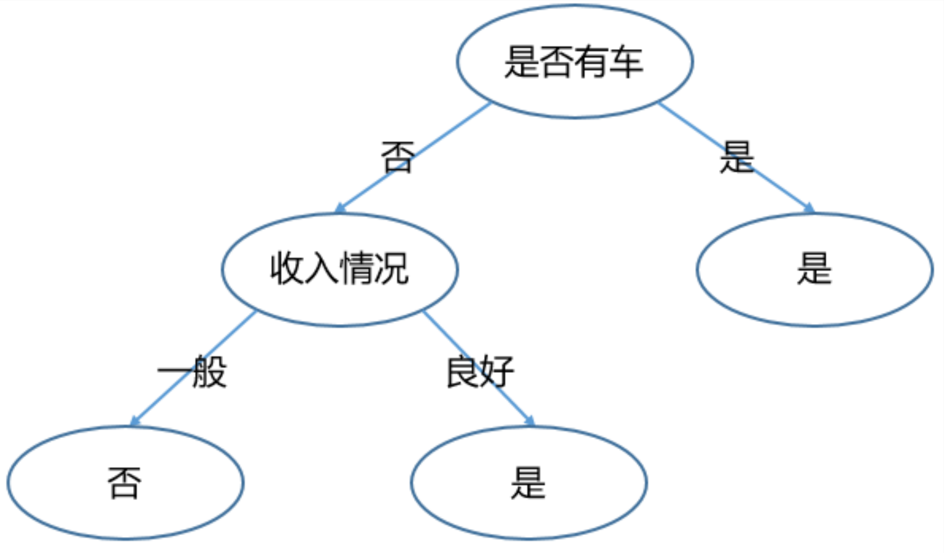
由以上对比可知，应该用“是否有车”作为划分属性，得到如下部分树：



第三章习题

解答：

1. 建树完成后得到的决策树如右图所示。



考虑对“收入情况”节点进行后剪枝，首先计算剪枝前验证集准确率为100%，不需要再计算剪枝后准确率，因为正确率已经达到最大值，故剪枝结束，最终的决策树如下图所示。

后剪枝注意事项：

1）列出所有非叶子节点后，从底向上考虑

2）如果剪枝后准确率小于等于剪枝前，则放弃剪枝。

第三章习题

解答：

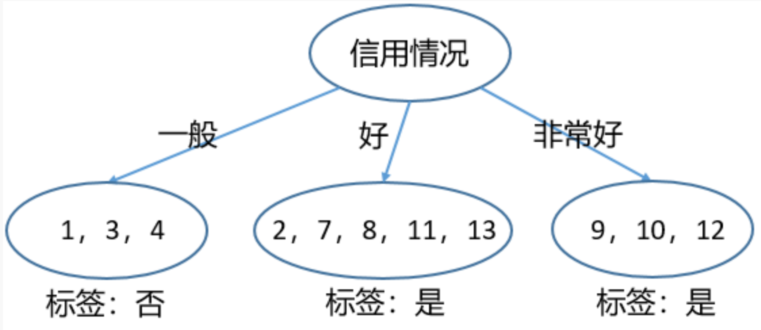
2. 首先计算各属性的基尼指数，如下：

$$GiniIndex(年龄)=\frac{4}{11}\left(1-\frac{3}{4}*\frac{3}{4}-\frac{1}{4}*\frac{1}{4}\right)+\frac{4}{11}*0+\frac{3}{11}\left(1-\frac{1}{3}*\frac{1}{3}-\frac{2}{3}*\frac{2}{3}\right)=0.258$$

$$GiniIndex(是否有车)=\frac{6}{11}\left(1-\frac{2}{6}*\frac{2}{6}-\frac{4}{6}*\frac{4}{6}\right)+\frac{5}{11}*0=0.242$$

$$GiniIndex(收入情况)=\frac{8}{11}\left(1-\frac{4}{8}*\frac{4}{8}-\frac{4}{8}*\frac{4}{8}\right)+\frac{3}{11}*0=0.364$$

$$GiniIndex(信用情况)=\frac{3}{11}*0+\frac{5}{11}*\left(1-\frac{1}{5}*\frac{1}{5}-\frac{4}{5}*\frac{4}{5}\right)+\frac{3}{11}*0=0.145$$



由以上对比可知，应选用“信用情况”作为划分属性，可得到如右图决策树。

考虑预剪枝，划分前：由于训练集根节点中正例较多，因此将根结点标签定为“是”，可计算出验证集准确率为50%；划分后，准确率为50%。划分前后验证集准确率没有变化，说明这次划分没有提升模型的泛化能力，所以停止划分。

预剪枝注意事项：

1）划分一次剪一次

2）如果划分后验证集准确率小于等于划分前，则放弃划分

第四章习题

选择题

- 神经网络必须具备的性质：非线性，可以通过反向传播训练

一、选择题

1. **（多选题）**下列针对激活函数（如 ReLU，Sigmoid，Tanh 等）的说法，正确的是（BD）

   A. 它们必须是可导的

   B. 它们可以在某些点不可导，但这些不可导的点必须很少

   C. 它们可以是任意的连续函数

   D. 它们必须是非线性的

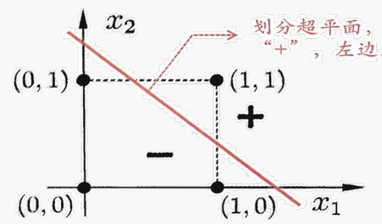
2. **（单选题）**下列哪类谓词逻辑问题不能用简单感知器解决？（D）

   A. 与

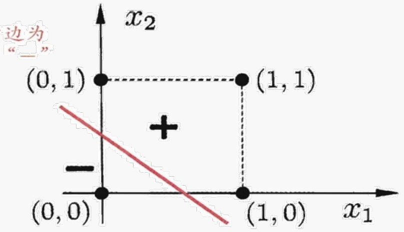
   B. 或

   C. 非

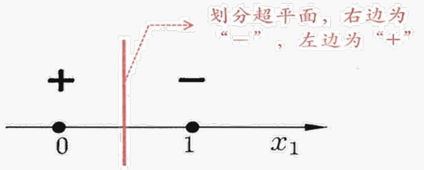
   D. 异或



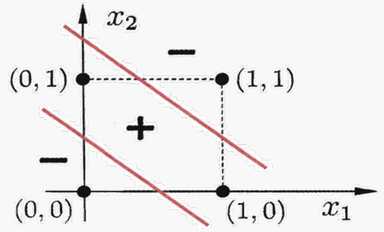
"与"运算



"或"运算



"非"运算



"异或"运算

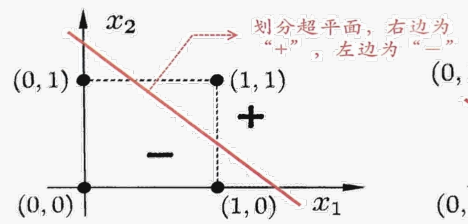
第四章习题

选择题

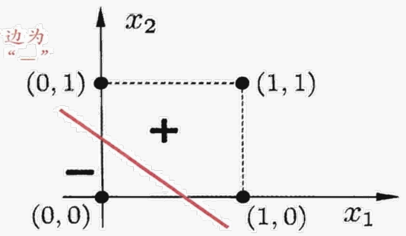
- 神经网络必须具备的性质：非线性，可以通过反向传播训练

一、选择题

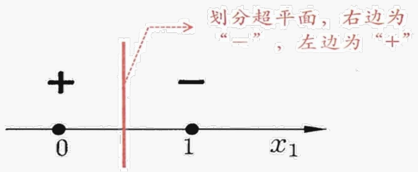
1. **（多选题）**下列针对激活函数（如 ReLU，Sigmoid，Tanh 等）的说法，正确的是（BD）
   - A. 它们必须是可导的
   - B. 它们可以在某些点不可导，但这些不可导的点必须很少
   - C. 它们可以是任意的连续函数
   - D. 它们必须是非线性的
2. **（单选题）**下列哪类谓词逻辑问题不能用简单感知器解决？（D）
   - A. 与
   - B. 或
   - C. 非
   - D. 异或



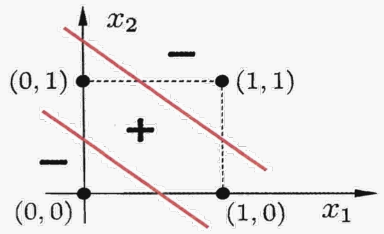
"与"运算



"或"运算



"非"运算



"异或"运算

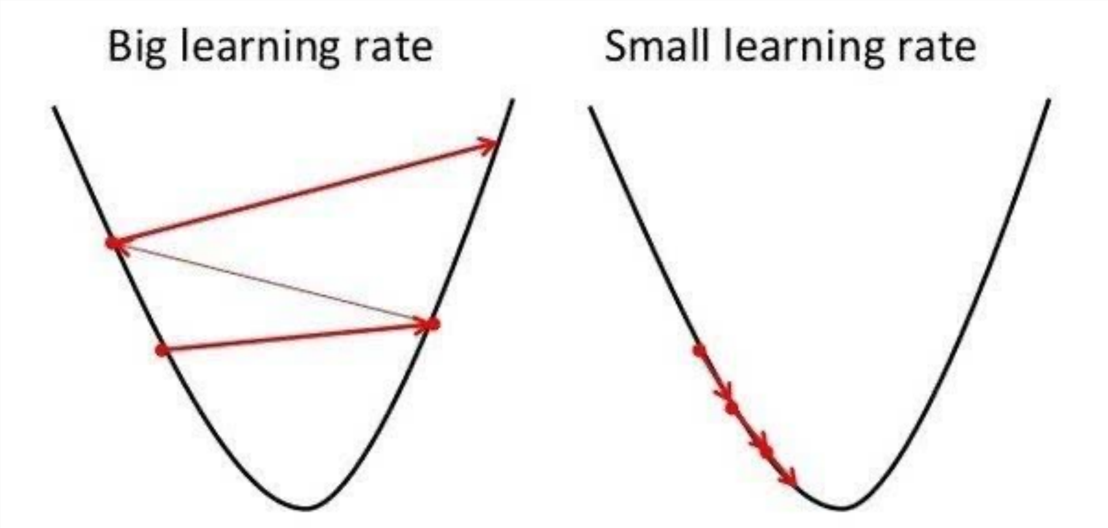
第四章习题

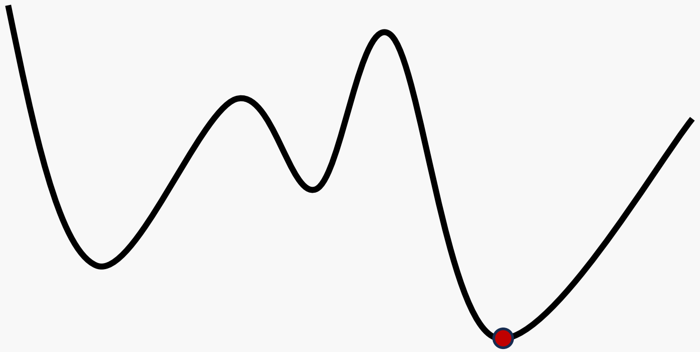
简答题

二、简答题

请分析采用反向传播算法训练神经网络时，学习率过大或过小时会产生什么负面影响。

- 学习率过大：训练振动，容易跳过极小值点
- 学习率过小：容易陷入局部最优；训练较慢





第四章习题

证明题

三、证明题

试证明：多层感知机中节点的激活函数如果采用线性函数，网络无法实现非线性映射。

- 感知机的运算本质就是矩阵的线性运算，如果激活函数也采用线性，最终运算结果也是输入的线性组合

第四章习题

计算题

四、计算题

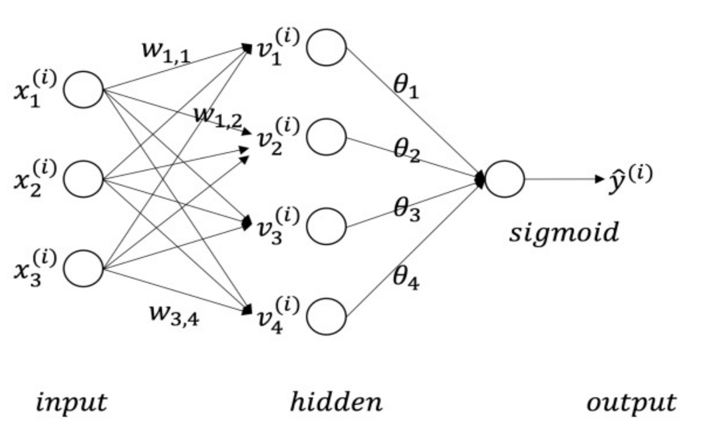
给定前向传播神经网络如下图所示，$\{\boldsymbol{x}^{(i)}, y^{(i)}\}, i = 1, \ldots, N$为样本，$\boldsymbol{x}^{(i)} \in \mathbb{R}^m$为输入。

$$v_j^{(i)} = \sigma\left(\sum_{k=1}^{3} w_{k,j}\, x_k^{(i)}\right), \qquad \sigma(z) = \frac{1}{1+exp(-z)}$$

输出$\hat{y}^{(i)} = \sigma(\boldsymbol{\theta}^T \boldsymbol{v}^{(i)})$，损失函数为

$$\mathcal{L} = -\frac{1}{N}\sum_{i=1}^{N}\left(y^{(i)} \log \hat{y}^{(i)} + (1-y^{(i)}) \log(1-\hat{y}^{(i)})\right)$$

1. 求$\frac{\partial \mathcal{L}}{\partial \theta_2}$。
2. 设学习率为η，求用随机梯度下降法更新$w_{1,2}$的表达式。



解答：

(1)

$$\frac{\partial \mathcal{L}}{\partial \hat{y}^{(i)}} = -\frac{1}{N}\left(\frac{y^{(i)}}{\hat{y}^{(i)}} + \frac{1-y^{(i)}}{\hat{y}^{(i)}-1}\right) = -\frac{1}{N}\frac{\hat{y}^{(i)}-y^{(i)}}{\hat{y}^{(i)}(\hat{y}^{(i)}-1)}$$

$$\frac{\partial \hat{y}^{(i)}}{\partial z^{(i)}} = \hat{y}^{(i)}(1-\hat{y}^{(i)})$$

$$\frac{\partial z^{(i)}}{\partial \theta_2} = v_2^{(i)}$$

则

$$\frac{\partial \mathcal{L}}{\partial \theta_2} = \sum_{i=1}^{N}\frac{\partial \mathcal{L}}{\partial \hat{y}^{(i)}}\frac{\partial \hat{y}^{(i)}}{\partial z^{(i)}}\frac{\partial z^{(i)}}{\partial \theta_2} = \frac{1}{N}\sum_{i=1}^{N}(\hat{y}^{(i)}-y^{(i)})v_2^{(i)}$$

(2)

$$\frac{\partial \hat{y}^{(i)}}{\partial v_2^{(i)}} = \hat{y}^{(i)}(1-\hat{y}^{(i)})\theta_2, \quad \frac{\partial v_2^{(i)}}{\partial w_{1,2}} = v_2^{(i)}\left(1-v_2^{(i)}\right)x_1^i$$

$$\frac{\partial \mathcal{L}}{\partial w_{1,2}} = \sum_{i=1}^{N}\frac{\partial \mathcal{L}}{\partial \hat{y}^{(i)}}\frac{\partial \hat{y}^{(i)}}{\partial v_2^{(i)}}\frac{\partial v_2^{(i)}}{\partial w_{1,2}} = -\frac{\hat{y}^{(i)}-y^{(i)}}{\hat{y}^{(i)}(\hat{y}^{(i)}-1)}\hat{y}^{(i)}(1-\hat{y}^{(i)})\theta_2 v_2^{(i)}\left(1-v_2^{(i)}\right)x_1^i$$

$$= (\hat{y}^{(i)}-y^{(i)})v_2^{(i)}\left(1-v_2^{(i)}\right)\theta_2 x_1$$

则，$w_{1,2} \leftarrow w_{1,2} - \eta(\hat{y}^{(i)}-y^{(i)})v_2^{(i)}\left(1-v_2^{(i)}\right)\theta_2 x_1$

第五章习题

一、选择题（3分）

1. （单选题）SVM的有效性取决于（ D ）

A. 核函数选择　B. 核函数参数　C. 软边距参数　D. 以上所有

2. （多选题）下列说法中错误的是（ BC ）

$$w^* = \sum_{i=1}^{n} \alpha_i^* y_i \boldsymbol{x}_i$$

$$\hat{y} = \mathrm{sgn}\left(\sum_{i=1}^{n} \alpha_i^* y_i (\boldsymbol{x}_i \cdot \boldsymbol{x}) + b^*\right)$$

A. 支持向量机的分类超平面只与支持向量有关

B. SVM是一种常见的线性分类器，它不能对数据做非线性分类

C. 软间隔支持向量机$\min \frac{1}{2}||w||^2 + C\sum_{i=1}^{N} \xi_i$中的C越大，模型的复杂度越低

D. 核函数能将数据特征提升到更高维空间

$$\begin{aligned} \min_{w,b,\xi} \quad & \frac{1}{2}\|\boldsymbol{w}\|_2^2 + C\sum_{i=1}^{n} \xi_i \\ \text{s.t.} \quad & y_i(\boldsymbol{w} \cdot \boldsymbol{x}_i + b) \ge 1 - \xi_i, \quad i = 1, \dots, n \\ & \xi_i \ge 0, \quad i = 1, \dots, n \end{aligned}$$

第五章习题

二、计算题

1. 给定正例$x_1=(1,1)$，$x_2=(-1,3)$，负例$x_3=(2,3), x_4=(2,4)$，计算支持向量、分类决策函数和分类超平面方程。（3分）

解法一：用对偶形式计算

- 首先写出对偶形式：

$$\max \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j} \alpha_i \alpha_j y_i y_j (x_i \cdot x_j)$$

$$s.t.\ \alpha_1 + \alpha_2 = \alpha_3 + \alpha_4$$

$$\alpha_i \geq 0$$

- 将$\alpha_4 = \alpha_1 + \alpha_2 - \alpha_3$代入上式得到目标函数

$$S(\alpha_1, \alpha_2, \alpha_3) = 2\alpha_1 + 2\alpha_2 - \frac{1}{2}(10\alpha_1^2 + 10\alpha_2^2 + \alpha_3^2 + 12\alpha_1\alpha_2 - 6\alpha_1\alpha_3 - 2\alpha_2\alpha_3)$$

第五章习题

二、计算题

解法一：用对偶形式计算

- 将目标函数对$\alpha_1, \alpha_2, \alpha_3$分别求偏导，令偏导数为0：

$$\frac{\partial S}{\partial \alpha 1} = 2 - 10\alpha 1 - 6\alpha 2 + 3\alpha 3 = 0$$

$$\frac{\partial S}{\partial \alpha 2} = 2 - 10\alpha 2 - 6\alpha 1 + \alpha 3 = 0$$

$$\frac{\partial S}{\partial \alpha 3} = -\alpha 3 + 3\alpha 1 + \alpha 2 = 0$$

- 方程组无解，说明最大值在边界处，$\alpha_1, \alpha_2, \alpha_3, \alpha_4$中存在0
- 当$\alpha_1 = 0$时，解得$\alpha_1 = 0, \alpha_2 = \alpha_3 = \frac{2}{9}, \alpha_4 = 0$，带入可得$S = -\frac{4}{9}$
- 当$\alpha_2 = 0$时，解得$\alpha_1 = \alpha_3 = \frac{2}{5}, \alpha_2 = \alpha_4 = 0$，代入得$S = \frac{2}{5}$
- 当$\alpha_3 = 0$时，解得$\alpha_1 = \alpha_2 = \frac{1}{8}, \alpha_4 = \frac{1}{4}$，带入可得$S = \frac{1}{4}$
- 当$\alpha_4 = 0$时，解得$\alpha_1 = \frac{1}{3}, \alpha_2 = \frac{1}{9}, \alpha_3 = \frac{4}{9}$，带入可得$S = \frac{4}{9}$

第五章习题

二、计算题

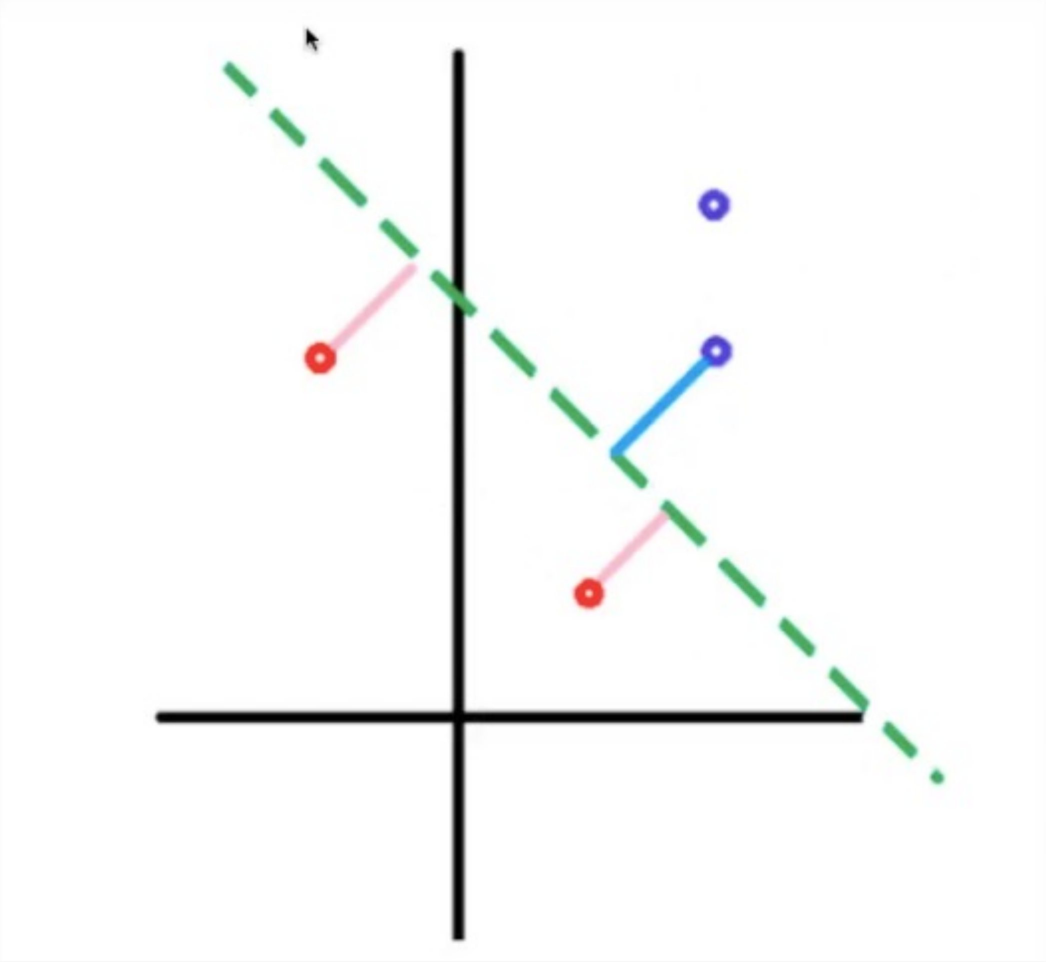
解法二：直接画图

- 画出各个样本的坐标
- 直接确定支持向量
- 非支持向量的α为0
- 计算得到超平面方程：

$$-\frac{2}{3}x1-\frac{2}{3}x2+\frac{7}{3}=0$$

- 分类决策函数：

$$\hat{y}=sgn(w^{*}x+b^{*})=sgn(-\frac{2}{3}x1-\frac{2}{3}x2+\frac{7}{3})$$



第五章习题

二、计算题

2. 考虑SVM（支持向量机）和来自两个类别的训练数据如下：

（1）通过拉格朗日法构建出SVM求解最优超平面的对偶优化问题。（1分）

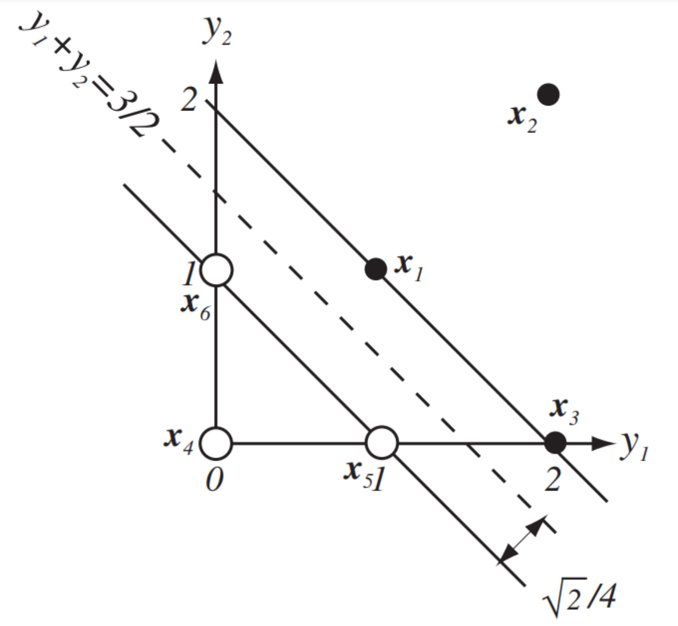
（2）求解支持向量。（2分）

（3）画出表格中6个数据点，支持向量，最优超平面以及决策边界。（1分）

支持向量机的对偶问题：

$$
\begin{aligned}
&\max_{\boldsymbol{\alpha}} \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j (\boldsymbol{x}_i \cdot \boldsymbol{x}_j) \\
&\text{s.t.} \quad \alpha_i \geq 0, \qquad i = 1,..,n \\
&\qquad \sum_{i=1}^{n} \alpha_i y_i = 0
\end{aligned}
$$

直接画图求解



第六章习题

一、选择题（3分）

1. （单选题）若将经验风险中的损失函数由$\lambda(\hat{\theta},\theta)=(\theta-\hat{\theta})^2$替换为$\lambda(\hat{\theta},\theta)=|\theta-\hat{\theta}|$，则$\theta$的贝叶斯估计量为（ B ）

A. 后验均值　　B. 后验中值　　C. 后验最大值　　D. 以上都不对

2. （单选题）关于参数估计，下面哪项说法不正确（ C ）

A. 当训练样本数无穷多的时候，最大似然估计和贝叶斯估计的结果是一样的

B. 最大似然估计比较简单直观

C. 贝叶斯估计就是后验分布的最大值点　后验均值

D. 若先验概率的信息是可靠的，那么可以认为贝叶斯估计比最大似然估计的结果更准确

$$\min_{\hat{\theta}}\int_{-\infty}^{\infty}|\theta-\hat{\theta}|p(\theta|D)d\theta=\int_{-\infty}^{\hat{\theta}}(\hat{\theta}-\theta)p(\theta|D)d\theta+\int_{\hat{\theta}}^{\infty}(\theta-\hat{\theta})p(\theta|D)d\theta \qquad \int_{-\infty}^{\hat{\theta}}p(\theta|D)d\theta-\int_{\hat{\theta}}^{\infty}p(\theta|D)d\theta=0$$

第六章习题

二、计算题

1. （贝叶斯决策）在二分类任务中，假定类型w_1和w_2分别为两个目标类别，已知先验概率分别为$P(w_1) = 0.2$和$P(w_2) = 0.8$，两类概率密度函数分别表示如下：

$$P(x|\omega_1) = \begin{cases} x, & 0 \le x < 1 \\ 2 - x, & 1 \le x \le 2 \\ 0, & else \end{cases} \qquad P(x|\omega_2) = \begin{cases} x - 1, & 1 \le x < 2 \\ 3 - x, & 2 \le x \le 3 \\ 0, & else \end{cases}$$

（1）求贝叶斯最小错误率准则下的判决域，并判断样本x=1.5属于哪一类；

（2）使用（1）中的判别准则和判别域，求总错误概率P(e)

解答：

（1）比较$P(\omega_1)P(x|\omega_1)$与$P(\omega_2)P(x|\omega_2)$可得到决策分类面x=1.2

判决域：当x<1.2时，分类为ω_1类，当x>1.2时，分类为ω_2类。所以x=1.5时，会被分为ω_2类

（2）$P(e) = \int_{-\infty}^{1.2} P(\omega_2)P(x|\omega_2)dx + \int_{1.2}^{+\infty} P(\omega_1)P(x|\omega_1)dx$=0.08

第六章习题

二、计算题

2. （最大似然估计）假设$(X_1, \ldots, X_n)$是来自概率密度函数如下的一组随机样本，求θ的最大似然估计。

$$f(x|\theta) = \theta x^{-2} \qquad 0 < \theta \leq x < \infty$$

解答：

$$L(\boldsymbol{x}; \theta) = \prod_{i=1}^{n} f(x_i|\theta) = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{-2}$$

$$l(\boldsymbol{x}; \theta) = \log L(\boldsymbol{x}; \theta) = n \log \theta - 2 \sum_{i=1}^{n} \log x_i$$

$$\frac{\partial l(\boldsymbol{x}; \theta)}{\partial \theta} = \frac{n}{\theta} > 0$$

则$l(\boldsymbol{x}; \theta)$单调递增。根据$0 < \theta \leq x < \infty$，则有

$$\hat{\theta} = \min\{X_1, \cdots, X_n\}$$