
机器学习1
第1章 绪论
基本术语
样本、属性
回归、分类
监督 、无监督
归纳偏好
奥卡姆剃刀
没有免费的午餐
发展历程
推理期、知识期、学习期
符号主义
连接主义
统计学习
第2章 模型评估
误差与过拟合
错误率、精度
过拟合（原因）
评估方法
留出法
交叉验证
留一法
性能度量
回归：均方误差
查准率、查全率、F1
ROC曲线、AUC
假设检验
偏差与方差
第3章 线性模型
线性回归
最小二乘法
多元线性回归：矩阵形式
广义线性模型
对数几率回归
单位阶跃函数
对数几率函数
对数似然
线性判别分析LDA
散度矩阵
目标函数
拉格朗日乘子法的求解
多分类学习
一对一OVO
一对其余OVR
第4章 决策树
基本流程
树的结构
递归过程
划分选择
纯度、信息熵
信息增益
增益率，属性偏好
基尼指数
剪枝处理
剪枝的目的
基本策略：预剪枝、后剪枝
连续与缺失值
连续值处理：划分点
缺失值处理：权重
选择指标的推广
第5章 神经网络
神经元模型
M-P神经元模型
激活函数：阶跃、sigmoid
感知机与多层网络
感知机结构
权重调整公式
异或问题-单隐层网络
误差逆传播算法
网络结构
前向过程计算
均方误差
梯度下降法与权值调整
链式求导与梯度反向传播
过拟合缓解策略：早停、正则化
全局最小与局部最小

机器学习2

- 第6章 支持向量机
  - 间隔与支持向量
    - 超平面与距离
    - 最大间隔、支持向量
    - SVM基本型
  - 对偶问题
    - 拉格朗日乘子法
    - 对偶问题形式
    - KKT条件与支持向量
  - 核函数
    - 高维特征空间映射
    - 常用核函数
  - 软间隔与正则化
    - 软间隔
    - 损失函数：0/1、hinge
    - 松弛变量
- 第7章 贝叶斯分类器
  - 贝叶斯决策论
    - 条件风险
    - 贝叶斯判定准则：贝叶斯定理
  - 参数估计
    - 极大似然估计
    - 朴素贝叶斯（属性条件独立性假设）
    - EM算法（处理隐变量）
- 第8章 集成学习
  - 个体与集成
    - 核心目标：结合多个弱学习器提升泛化性能
    - 基学习器：好而不同（多样性、准确性）
  - 集成学习方法
    - Boosting（串行，AdaBoost：样本权重调整）
    - Bagging（并行，自助采样）
    - 随机森林（属性随机选择）
  - 结合策略
    - 平均法
    - 投票法
    - 学习法
- 第9章 聚类
  - 聚类任务
    - 无监督学习：样本无标记
    - 聚类：簇
  - 性能度量
    - 外部指标
    - 内部指标
  - 距离计算
    - 闵可夫斯基距离
    - 无序属性的VDM
  - 原型聚类
    - k-均值算法
    - 学习向量量化
    - 高斯混合聚类（EM算法）
  - 密度聚类：DBSCAN
  - 层次聚类：AGNES（自底向上合并）
- 第10章 降维
  - k近邻学习
  - 低维嵌入
    - 维数灾难
    - 多维缩放（MDS）
  - 主成分分析
    - 最近重构性
    - 最大可分性
    - 协方差矩阵特征值分解
  - 非线性降维
    - 核化线性降维（KPCA）
    - 流形学习
      - Isomap
      - LLE
- 第11章 特征选择
  - 子集搜索与评价
    - 前向/后向/双向
    - 信息增益
  - 特征选择
    - 过滤式（Relief）
    - 包裹式（LVW）
    - 嵌入式（L1正则化）

2.2

- 10折交叉检验：每次训练样本中正反例数目一样（各90个或者概率一样），按算法规则，随机猜测，测试样本判断为正反例的概率是一样的，所以错误率的期望是50%。
- 留一法：如果留下的是正例，训练样本中反例的数目比正例多一个，按算法规则留下的样本会被判断是反例；同理，留出的是反例，则会被判断成正例，所以错误率是100%。

$$
\begin{aligned}
\frac{\partial E_{(w,b)}}{\partial w} &= \frac{\partial}{\partial w}\left[\sum_{i=1}^{m}(y_i - wx_i - b)^2\right] \\
&= \sum_{i=1}^{m}\frac{\partial}{\partial w}\left[(y_i - wx_i - b)^2\right] \\
&= \sum_{i=1}^{m}\left[2\cdot(y_i - wx_i - b)\cdot(-x_i)\right] \\
&= \sum_{i=1}^{m}\left[2\cdot\left(wx_i^2 - y_ix_i + bx_i\right)\right] \\
&= 2\cdot\left(w\sum_{i=1}^{m}x_i^2 - \sum_{i=1}^{m}y_ix_i + b\sum_{i=1}^{m}x_i\right) \\
&= 2\left(w\sum_{i=1}^{m}x_i^2 - \sum_{i=1}^{m}(y_i - b)\,x_i\right)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial E_{(w,b)}}{\partial b} &= \frac{\partial}{\partial b}\left[\sum_{i=1}^{m}(y_i - wx_i - b)^2\right] \\
&= \sum_{i=1}^{m}\frac{\partial}{\partial b}\left[(y_i - wx_i - b)^2\right] \\
&= \sum_{i=1}^{m}\left[2\cdot(y_i - wx_i - b)\cdot(-1)\right] \\
&= \sum_{i=1}^{m}\left[2\cdot(b - y_i + wx_i)\right] \\
&= 2\cdot\left[\sum_{i=1}^{m}b - \sum_{i=1}^{m}y_i + \sum_{i=1}^{m}wx_i\right] \\
&= 2\left(mb - \sum_{i=1}^{m}(y_i - wx_i)\right)
\end{aligned}
$$

[推导]：令公式 (3.5) 等于 0

$$0 = w\sum_{i=1}^{m}x_i^2 - \sum_{i=1}^{m}(y_i - b)x_i$$

$$w\sum_{i=1}^{m}x_i^2 = \sum_{i=1}^{m}y_ix_i - \sum_{i=1}^{m}bx_i$$

由于令公式 (3.6) 等于 0 可得 $b = \frac{1}{m}\sum_{i=1}^{m}(y_i - wx_i)$，又因为 $\frac{1}{m}\sum_{i=1}^{m}y_i = \bar{y}$，$\frac{1}{m}\sum_{i=1}^{m}x_i = \bar{x}$，则 $b = \bar{y} - w\bar{x}$，代入上式可得

$$w\sum_{i=1}^{m}x_i^2 = \sum_{i=1}^{m}y_ix_i - \sum_{i=1}^{m}(\bar{y} - w\bar{x})x_i$$

$$w\sum_{i=1}^{m}x_i^2 = \sum_{i=1}^{m}y_ix_i - \bar{y}\sum_{i=1}^{m}x_i + w\bar{x}\sum_{i=1}^{m}x_i$$

$$w(\sum_{i=1}^{m}x_i^2 - \bar{x}\sum_{i=1}^{m}x_i) = \sum_{i=1}^{m}y_ix_i - \bar{y}\sum_{i=1}^{m}x_i$$

$$w = \frac{\sum_{i=1}^{m}y_ix_i - \bar{y}\sum_{i=1}^{m}x_i}{\sum_{i=1}^{m}x_i^2 - \bar{x}\sum_{i=1}^{m}x_i}$$

由于 $\bar{y}\sum_{i=1}^{m}x_i = \frac{1}{m}\sum_{i=1}^{m}y_i\sum_{i=1}^{m}x_i = \bar{x}\sum_{i=1}^{m}y_i$，$\bar{x}\sum_{i=1}^{m}x_i = \frac{1}{m}\sum_{i=1}^{m}x_i\sum_{i=1}^{m}x_i = \frac{1}{m}(\sum_{i=1}^{m}x_i)^2$，代入上式即可得公式 (3.7)

$$w = \frac{\sum_{i=1}^{m}y_i(x_i - \bar{x})}{\sum_{i=1}^{m}x_i^2 - \frac{1}{m}(\sum_{i=1}^{m}x_i)^2}$$

$D(x,y)=\{(2,3.2),(4,5.0),(6,6.8)\}$.

即求 $f(x_i)=wx_i+b$ 使 $f(x_i)\simeq y_i$

均方误差最小化：$(w^*,b^*)=\underset{(w,b)}{\arg\min}\sum_{i=1}^{m}(y_i-wx_i-b)^2$

$$E(w,b)=\sum_{i=1}^{m}(y_i-wx_i-b)^2$$

$$\begin{cases}\dfrac{\partial E}{\partial w}=2\left(w\sum_{i=1}^{m}x_i^2-\sum_{i=1}^{m}(y_i-b)x_i\right)=0\\[2ex]\dfrac{\partial E}{\partial b}=2\left(mb-\sum_{i=1}^{m}(y_i-wx_i)\right)=0\end{cases}$$

$$\Rightarrow\begin{cases}w=\dfrac{\sum_{i=1}^{m}y_i(x_i-\bar{x})}{\sum_{i=1}^{m}x_i^2-\frac{1}{m}\left(\sum_{i=1}^{m}x_i\right)^2}\\[4ex]b=\dfrac{1}{m}\sum_{i=1}^{m}(y_i-wx_i)\end{cases}$$

代入 $m=3$，$\bar{x}=\dfrac{2+4+6}{3}=4$

$$\Rightarrow\begin{cases}w=0.9\\b=1.4\end{cases}\Rightarrow f(x)=0.9x+1.4$$

[推导]：由公式 (3.36) 可得拉格朗日函数为

$$L(\boldsymbol{w},\lambda)=-\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_b\boldsymbol{w}+\lambda(\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_w\boldsymbol{w}-1)$$

对 $\boldsymbol{w}$ 求偏导可得

$$\begin{aligned}\frac{\partial L(\boldsymbol{w},\lambda)}{\partial\boldsymbol{w}}&=-\frac{\partial(\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_b\boldsymbol{w})}{\partial\boldsymbol{w}}+\lambda\frac{\partial(\boldsymbol{w}^{\mathrm{T}}\mathbf{S}_w\boldsymbol{w}-1)}{\partial\boldsymbol{w}}\\&=-(\mathbf{S}_b+\mathbf{S}_b^{\mathrm{T}})\boldsymbol{w}+\lambda(\mathbf{S}_w+\mathbf{S}_w^{\mathrm{T}})\boldsymbol{w}\end{aligned}$$

由于 $\mathbf{S}_b=\mathbf{S}_b^{\mathrm{T}},\mathbf{S}_w=\mathbf{S}_w^{\mathrm{T}}$，所以

$$\frac{\partial L(\boldsymbol{w},\lambda)}{\partial\boldsymbol{w}}=-2\mathbf{S}_b\boldsymbol{w}+2\lambda\mathbf{S}_w\boldsymbol{w}$$

令上式等于 0 即可得

$$-2\mathbf{S}_b\boldsymbol{w}+2\lambda\mathbf{S}_w\boldsymbol{w}=0$$

$$\mathbf{S}_b\boldsymbol{w}=\lambda\mathbf{S}_w\boldsymbol{w}$$

2.解 两类样本集 $w_1=\{(1,0),(1,1),(2,0)\}$

$w_2=\{(-1,1),(-1,0),(0,1)\}$

设计LDA分类器：使同类样例的投影点尽可能接近

使异类样例的投影点尽可能远离

则要使J最大化，$J=\dfrac{\|w^T\mu_0-w^T\mu_1\|_2^2}{w^T\Sigma_0 w+w^T\Sigma_1 w}$

$=\dfrac{w^T(\mu_0-\mu_1)(\mu_0-\mu_1)^T w}{w^T(\Sigma_0+\Sigma_1)w}$

类内散度矩阵 $S_w=\Sigma_0+\Sigma_1=\sum_{x\in X_0}(x-\mu_0)(x-\mu_0)^T+\sum_{x\in X_1}(x-\mu_1)(x-\mu_1)^T$

类间散度矩阵 $S_b=(\mu_0-\mu_1)(\mu_0-\mu_1)^T$

则 $J=\dfrac{w^TS_bw}{w^TS_ww}$ 令 $w^TS_ww=1$.

即求 $\min\limits_{w} -w^TS_bw$

s.t. $w^TS_ww=1$.

拉格朗日乘子法 $\vec{w}=S_w^{-1}(\vec{\mu_0}-\vec{\mu_1})$

由题目数据得 $\mu_0=(\frac{4}{3},\frac{1}{3})^T$ $\mu_1=(-\frac{2}{3},\frac{2}{3})^T$

$S_w=\begin{pmatrix}-\frac{1}{3}\\-\frac{1}{3}\end{pmatrix}\begin{pmatrix}-\frac{1}{3} & -\frac{1}{3}\end{pmatrix}+\begin{pmatrix}-\frac{1}{3}\\\frac{2}{3}\end{pmatrix}\begin{pmatrix}-\frac{1}{3} & \frac{2}{3}\end{pmatrix}+\begin{pmatrix}\frac{2}{3}\\-\frac{1}{3}\end{pmatrix}\begin{pmatrix}\frac{2}{3} & -\frac{1}{3}\end{pmatrix}$

$+\begin{pmatrix}-\frac{1}{3}\\\frac{1}{3}\end{pmatrix}\begin{pmatrix}-\frac{1}{3} & \frac{1}{3}\end{pmatrix}+\begin{pmatrix}-\frac{1}{3}\\-\frac{2}{3}\end{pmatrix}\begin{pmatrix}-\frac{1}{3} & -\frac{2}{3}\end{pmatrix}+\begin{pmatrix}\frac{2}{3}\\\frac{1}{3}\end{pmatrix}\begin{pmatrix}\frac{2}{3} & \frac{1}{3}\end{pmatrix}$

$=\begin{pmatrix}\frac{4}{3} & 0\\0 & \frac{4}{3}\end{pmatrix}$

$\therefore \vec{w}=S_w^{-1}(\mu_0-\mu_1)=\begin{pmatrix}\frac{3}{4} & 0\\0 & \frac{3}{4}\end{pmatrix}\cdot\begin{pmatrix}2\\-\frac{1}{3}\end{pmatrix}=\begin{pmatrix}\frac{3}{2}\\-\frac{1}{4}\end{pmatrix}$

将 $(0,-1)^T$ 投影得 $y=\vec{w}^T\begin{pmatrix}0\\-1\end{pmatrix}=\frac{1}{4}$

$\vec{w}^T\vec{\mu_0}=\begin{pmatrix}\frac{3}{2} & -\frac{1}{4}\end{pmatrix}\begin{pmatrix}\frac{4}{3}\\\frac{1}{3}\end{pmatrix}=\frac{23}{12}$

$\vec{w}^T\vec{\mu_1}=\begin{pmatrix}\frac{3}{2} & -\frac{1}{4}\end{pmatrix}\begin{pmatrix}-\frac{2}{3}\\\frac{2}{3}\end{pmatrix}=-\frac{7}{6}$

$(\vec{w}^T\vec{\mu_0}+\vec{w}^T\vec{\mu_1})/2=\frac{3}{8}>\frac{1}{4}$ $\therefore (0,-1)$ 为 w_2 类.

表 5.1　贷款申请样本数据表

ID	年龄	有工作	有自己的房子	信贷情况	类别
1	青年	否	否	一般	否
2	青年	否	否	好	否
3	青年	是	否	好	是
4	青年	是	是	一般	是
5	青年	否	否	一般	否
6	中年	否	否	一般	否
7	中年	否	否	好	否
8	中年	是	是	好	是
9	中年	否	是	非常好	是
10	中年	否	是	非常好	是
11	老年	否	是	非常好	是
12	老年	否	是	好	是
13	老年	是	否	好	是
14	老年	是	否	非常好	是
15	老年	否	否	一般	否

决策树

首先计算经验熵 $H(D)$.

$$H(D)=-\frac{9}{15}\log_2\frac{9}{15}-\frac{6}{15}\log_2\frac{6}{15}=0.971$$

然后计算各特征对数据集 D 的信息增益. 分别以 A_1，A_2，A_3，A_4 表示年龄、有工作、有自己的房子和信贷情况 4 个特征，则

（1）

$$\begin{aligned}g(D,A_1)&=H(D)-\left[\frac{5}{15}H(D_1)+\frac{5}{15}H(D_2)+\frac{5}{15}H(D_3)\right]\\&=0.971-\left[\frac{5}{15}\left(-\frac{2}{5}\log_2\frac{2}{5}-\frac{3}{5}\log_2\frac{3}{5}\right)\right.\\&\quad\left.+\frac{5}{15}\left(-\frac{3}{5}\log_2\frac{3}{5}-\frac{2}{5}\log_2\frac{2}{5}\right)+\frac{5}{15}\left(-\frac{4}{5}\log_2\frac{4}{5}-\frac{1}{5}\log_2\frac{1}{5}\right)\right]\\&=0.971-0.888=0.083\end{aligned}$$

这里 D_1，D_2，D_3 分别是 D 中 A_1（年龄）取值为青年、中年和老年的样本子集. 类似地，

（2）

$$g(D,A_2)=H(D)-\left[\frac{5}{15}H(D_1)+\frac{10}{15}H(D_2)\right]$$

$$=0.971-\left[\frac{5}{15}\times 0+\frac{10}{15}\left(-\frac{4}{10}\log_2\frac{4}{10}-\frac{6}{10}\log_2\frac{6}{10}\right)\right]=0.324$$

（3）

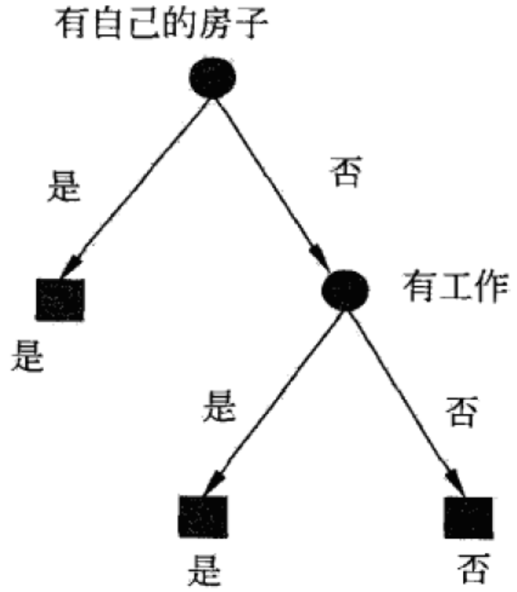
$$g(D,A_3)=0.971-\left[\frac{6}{15}\times 0+\frac{9}{15}\left(-\frac{3}{9}\log_2\frac{3}{9}-\frac{6}{9}\log_2\frac{6}{9}\right)\right]$$

$$=0.971-0.551=0.420$$

（4）

$$g(D,A_4)=0.971-0.608=0.363$$

最后，比较各特征的信息增益值．由于特征 A_3（有自己的房子）的信息增益值最大，所以选择特征 A_3 作为最优特征．■



对 D_2 则需从特征 A_1（年龄），A_2（有工作）和 A_4（信贷情况）中选择新的特征．计算各个特征的信息增益：

$$g(D_2,A_1)=H(D_2)-H(D_2\,|\,A_1)=0.918-0.667=0.251$$
$$g(D_2,A_2)=H(D_2)-H(D_2\,|\,A_2)=0.918$$
$$g(D_2,A_4)=H(D_2)-H(D_2\,|\,A_4)=0.474$$

$$
\begin{aligned}
\frac{\partial E_k}{\partial \theta_j} &= \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \theta_j} \\
&= \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial [f(\beta_j - \theta_j)]}{\partial \theta_j} \\
&= \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot f'(\beta_j - \theta_j) \times (-1) \\
&= \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot f(\beta_j - \theta_j) \times [1 - f(\beta_j - \theta_j)] \times (-1) \\
&= \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \hat{y}_j^k \left(1 - \hat{y}_j^k\right) \times (-1)
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\partial \left[\frac{1}{2} \sum_{j=1}^{l} \left(\hat{y}_j^k - y_j^k\right)^2 \right]}{\partial \hat{y}_j^k} \cdot \hat{y}_j^k \left(1 - \hat{y}_j^k\right) \times (-1) \\
&= \frac{1}{2} \times 2(\hat{y}_j^k - y_j^k) \times 1 \cdot \hat{y}_j^k \left(1 - \hat{y}_j^k\right) \times (-1) \\
&= (y_j^k - \hat{y}_j^k)\hat{y}_j^k \left(1 - \hat{y}_j^k\right) \\
&= g_j
\end{aligned}
$$

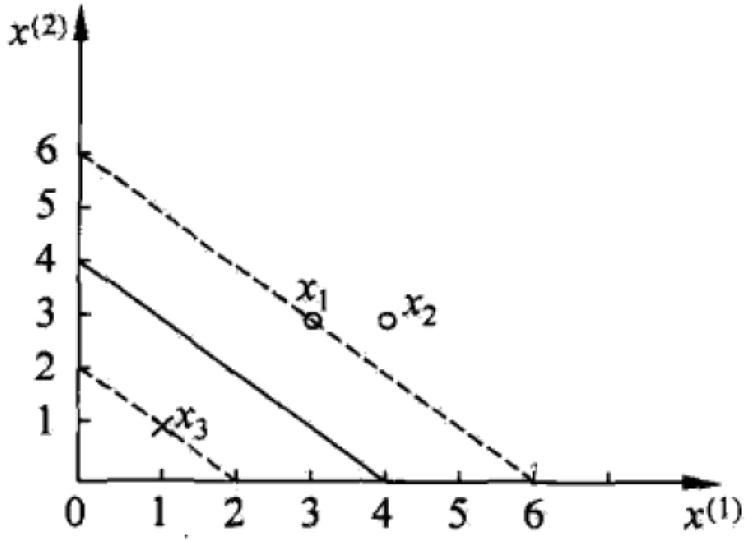
$$
\Delta \theta_j = -\eta \frac{\partial E_k}{\partial \theta_j} = -\eta g_j
$$

$$\Delta v_{ih} = -\eta \frac{\partial E_k}{\partial v_{ih}}$$

$$\begin{aligned}
\frac{\partial E_k}{\partial v_{ih}} &= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot \frac{\partial b_h}{\partial \alpha_h} \cdot \frac{\partial \alpha_h}{\partial v_{ih}} \\
&= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot \frac{\partial b_h}{\partial \alpha_h} \cdot x_i \\
&= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot \frac{\partial \beta_j}{\partial b_h} \cdot f'(\alpha_h - \gamma_h) \cdot x_i \\
&= \sum_{j=1}^{l} \frac{\partial E_k}{\partial \hat{y}_j^k} \cdot \frac{\partial \hat{y}_j^k}{\partial \beta_j} \cdot w_{hj} \cdot f'(\alpha_h - \gamma_h) \cdot x_i
\end{aligned}$$

$$\begin{aligned}
&= \sum_{j=1}^{l} (-g_j) \cdot w_{hj} \cdot f'(\alpha_h - \gamma_h) \cdot x_i \\
&= -f'(\alpha_h - \gamma_h) \cdot \sum_{j=1}^{l} g_j \cdot w_{hj} \cdot x_i \\
&= -b_h(1 - b_h) \cdot \sum_{j=1}^{l} g_j \cdot w_{hj} \cdot x_i \\
&= -e_h \cdot x_i
\end{aligned}$$

$$\Delta v_{ih} = -\eta \frac{\partial E_k}{\partial v_{ih}} = \eta e_h x_i$$



$$\underset{\boldsymbol{w},b}{\arg\min} \quad \frac{1}{2}\|\boldsymbol{w}\|^2$$
$$\text{s.t.} \quad y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) \geq 1, \; i = 1, 2, \ldots, m.$$

$$L(\boldsymbol{w}, b, \boldsymbol{\alpha}) = \frac{1}{2}\|\boldsymbol{w}\|^2 - \sum_{i=1}^{m} \alpha_i \left(y_i(\boldsymbol{w}^\top \boldsymbol{x}_i + b) - 1 \right)$$

$$\max_{\alpha} \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^T \boldsymbol{x}_j$$
$$s.t. \; \sum_{i=1}^{m} \alpha_i y_i = 0, \; \alpha_i \geq 0, \; i = 1, 2, \cdots, m.$$

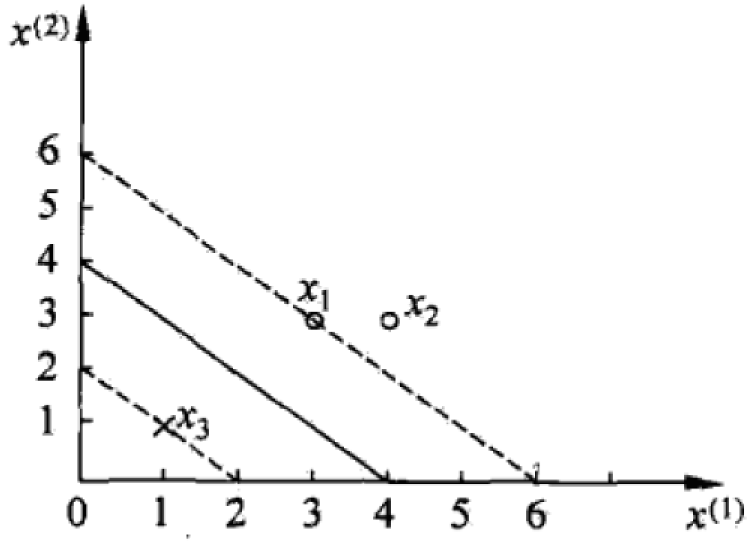
$$\min_{\boldsymbol{\alpha}} \quad \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^\top \boldsymbol{x}_j - \sum_{i=1}^{m} \alpha_i$$
$$\text{s.t.} \quad \sum_{i=1}^{m} \alpha_i y_i = 0, \; \alpha_i \geq 0, \; i = 1, 2, \ldots, m.$$

$$\min_{\alpha} \quad \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{N}\alpha_i$$

$$= \frac{1}{2}(18\alpha_1^2 + 25\alpha_2^2 + 2\alpha_3^2 + 42\alpha_1\alpha_2 - 12\alpha_1\alpha_3 - 14\alpha_2\alpha_3) - \alpha_1 - \alpha_2 - \alpha_3$$

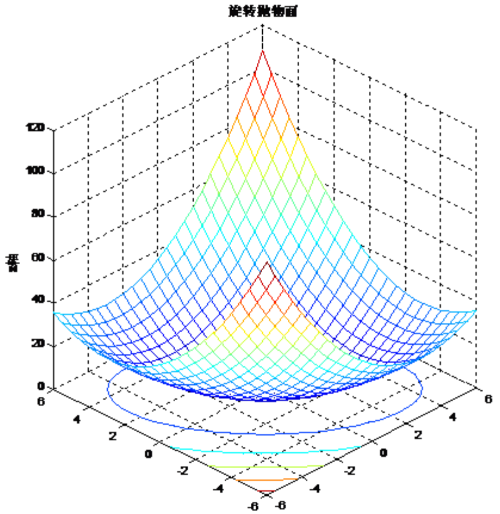
$$\text{s.t.} \quad \alpha_1 + \alpha_2 - \alpha_3 = 0$$

$$\alpha_i \geqslant 0, \quad i = 1,2,3$$



将 $\alpha_3 = \alpha_1 + \alpha_2$ 代入目标函数

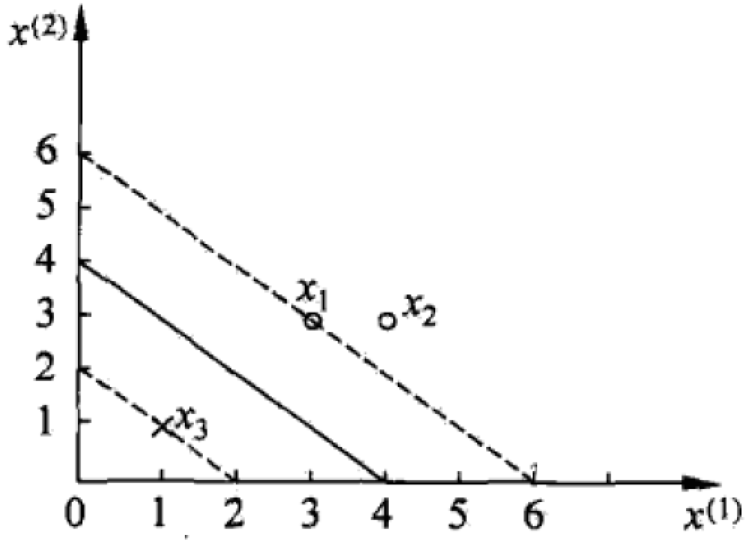
$$s(\alpha_1, \alpha_2) = 4\alpha_1^2 + \frac{13}{2}\alpha_2^2 + 10\alpha_1\alpha_2 - 2\alpha_1 - 2\alpha_2$$



对 α_1, α_2 求偏导数并令其为 0，易知 $s(\alpha_1, \alpha_2)$ 在点 $\left(\frac{3}{2}, -1\right)^{\mathrm{T}}$ 取极值，但该点不满足约束条件 $\alpha_2 \geqslant 0$，所以最小值应在边界上达到.

当 $\alpha_1 = 0$ 时，最小值 $s\left(0, \frac{2}{13}\right) = -\frac{2}{13}$；当 $\alpha_2 = 0$ 时，最小值 $s\left(\frac{1}{4}, 0\right) = -\frac{1}{4}$. 于是 $s(\alpha_1, \alpha_2)$ 在 $\alpha_1 = \frac{1}{4}, \alpha_2 = 0$ 达到最小，此时 $\alpha_3 = \alpha_1 + \alpha_2 = \frac{1}{4}$.

$\alpha_1^* = \alpha_3^* = \frac{1}{4}$ 对应的实例点 x_1, x_3 是支持向量



$$w_1^* = w_2^* = \frac{1}{2}$$

$$b^* = -2$$

分离超平面为

$$\frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)} - 2 = 0$$

分类决策函数为

$$f(x) = \operatorname{sign}\left(\frac{1}{2}x^{(1)} + \frac{1}{2}x^{(2)} - 2\right)$$

7.1

对于根蒂

	蜷缩	稍蜷	硬挺
好瓜	5	3	0
坏瓜	3	4	2

对于好瓜

$LL(\theta_{好瓜}) = 5\log\lambda_1 + 3\log\lambda_2$

$\frac{d(LL(\theta_{好瓜}))}{d\lambda_1} = \frac{5}{\lambda_1}$ $\lambda_1 = \frac{5}{8}, \lambda_2 = \frac{3}{8}$

$\frac{d(LL(\theta_{好瓜}))}{d\lambda_2} = \frac{3}{\lambda_2}$

对于坏瓜

$LL(\theta_{坏瓜}) = 3\log\lambda_1' + 4\log\lambda_2' + 2\log(1-\lambda_1'-\lambda_2')$

求偏导取零得

$\lambda_1 = \frac{3}{9}$ $\lambda_2 = \frac{4}{9}$ $1-\lambda_1-\lambda_2 = \frac{2}{9}$

对于敲声

	浊响	沉闷	清脆
好瓜	6	2	0
坏瓜	4	3	2

同理得 $\lambda_1 = \frac{6}{8}$ $\lambda_2 = \frac{2}{8}$ $1-\lambda_1-\lambda_2 = 0$

$\lambda_1' = \frac{4}{9}$ $\lambda_2' = \frac{3}{9}$ $1-\lambda_1'-\lambda_2' = \frac{2}{9}$

综上：

$P(色泽=青绿|好瓜=是) = \frac{3}{8}$

$P(色泽=乌黑|好瓜=是) = \frac{1}{2}$

$P(色泽=浅白|好瓜=是) = \frac{1}{8}$

$P(色泽=青绿|好瓜=否) = \frac{1}{3}$

$P(色泽=乌黑|好瓜=否) = \frac{2}{9}$

$P(色泽=浅白|好瓜=否) = \frac{4}{9}$

$P(根蒂=蜷缩|好瓜=是) = \frac{5}{8}$

$P(根蒂=稍蜷|好瓜=是) = \frac{3}{8}$

$P(根蒂=硬挺|好瓜=是) = 0$

$P(根蒂=蜷缩|好瓜=否) = \frac{1}{3}$

$P(根蒂=稍蜷|好瓜=否) = \frac{4}{9}$

$P(根蒂=硬挺|好瓜=否) = \frac{2}{9}$

$P(敲声=浊响|好瓜=是) = \frac{3}{4}$

$P(敲声=沉闷|好瓜=是) = \frac{1}{4}$

$P(敲声=清脆|好瓜=是) = 0$

$P(敲声=浊响|好瓜=否) = \frac{4}{9}$

$P(敲声=沉闷|好瓜=否) = \frac{1}{3}$

$P(敲声=清脆|好瓜=否) = \frac{2}{9}$