课时内容

第6章 机器学习

6 粗糙集的基本理论

◆ 信息系统

信息系统: 信息系统是RS理论研究的主体，它被抽象为一个数据集，即一张数据表

信息系统是一个四元组

IS=(U, A, V, f)

其中，U是对象的有限非空集，亦称域；A为属性的有限非空集；V是属性的值域； f是映射函数，即f：U×A→V。属性集A又可分为条件属性C和决策属性D两部分，即A=C∪D，且C∩D=Φ。上述数据表也称为决策表。

U \ A	C			D
	a_1	a_2	a_3	a_4
u_1	2	2	2	2
u_2	2	1	1	2
u_3	2	1	1	2
u_4	1	2	2	1
u_5	2	2	2	1
u_6	1	1	2	1

表中：U={u_1, u_2, u_3, u_4, u_5, u_6}； A={a_1 , a_2 , a_3 , a_4}；V={2, 1}；函数实现映射，例$f_{a1}(u_1)=2$；并且C={a_1 , a_2 , a_3 }, D={a_4 }

不分明关系

不分明关系作为RS理论的基础，亦称为等价关系

定义6.1 对信息系统IS={ U, A, V, f }，设B⊆A，不分明关系定义为

$$IND(B)=\{(x, y) \in U \times U \mid f_b(x) = f_b(y), \text{ for } b \in B\}$$

显然，若(x, y)∈IND(B)，则对象x, y在属性集B上是等价的，即不可区分的。

等价类和等价划分

1. 等价类

依据IND(B)，能将U划成不同类，可建立任意对象x关于B的**等价类** [x]B

定义6.2 设B ⊆ A，对所有x∈ U ，关于B的等价类[x]B定义为

$$[x]_B=\{ y \in U \mid (x, y) \in IND(B) \}$$

2. 等价划分

依据IND(B)，可将U划分为若干个等价类，这些等价类的集合又被称为**等价划分**，记为U/IND(B)，或简记为U/B 。

◆ 等价类和等价划分

例6.4 如下决策表所述信息系统，求其等价类和等价划分。

解：由其条件属性集C ={a1 , a2 , a3 },的IND(C), 可导出以下4个等价类：

[u1]C={u1 , u5}, [u2]C={u2 , u3}, [u4]C={u4}, [u6]C={u6}

从而有如下等价划分：

U/C={ {u1 , u5}, {u2 , u3}, {u4}, {u6} }

U \ A	C			D
	a_1	a_2	a_3	a_4
u_1	2	2	2	2
u_2	2	1	1	2
u_3	2	1	1	2
u_4	1	2	2	1
u_5	2	2	2	1
u_6	1	1	2	1

◆ 上近似和下近似

粗集的概念是通过上近似和下近似所确定的边界区域进行定义的。

定义6.3 设X ⊆ U, B ⊆ A, X对B的**下近似$B_-(X)$** 可定义为X所包含的关于B的所有等价类的并集：

$B_-(X) = \cup\{ [x]B \mid [x]B \subseteq X \}$，也即$B_-(X)=\{ x\in U \mid [x]B \subseteq X \}$

定义6.4 设X ⊆ U, B ⊆ A, X对B的**上近似$B^-(X)$**可定义为与X交集非空的关于B的所有等价类的并集：

$B^-(X)=\cup\{ [x]B \mid [x]B\cap X\neq\Phi\}$，也即$B^-(X)=\{ x\in U \mid [x]B\cap X\neq\Phi\}$

◆ 上近似和下近似

例6.5 对例6.4所述信息系统，令X= = {u1 , u2 ,u3 },求关于条件属性集C的上近似和下近似。

解：由例6.4已求出其条件属性集C ={a1 , a2 , a3 }的不分明关系IND(C), 可导出以下4个等价类：{u1 , u5}, {u2 , u3}, {u4}, {u6}

及其等价划分：U/C={ {u1 , u5}, {u2 , u3}, {u4}, {u6} }

由于可为X包含的等价类[u]C仅有{u2 , u3}, 即下近似为$\underline{C}(X)$ = {u2 , u3}

又由于{u1, u5}∩{u1, u2 ,u3 } ≠Φ, {u2, u3}∩{u1 , u2 ,u3 } ≠Φ，即上近似为

$\overline{C}(X)$ = {u1 , u2 , u3 , u5}

◆ 边界区域和粗集

定义6.5 设$X \subseteq U, B \subseteq A$，对象集X关于属性集B的**边界区域**定义为

$$BN_B(x) = B^-(X) - B_-(X)$$

定义6.6 设$X \subseteq U, B \subseteq A$，由对象集X关于属性集B的边界区域的定义，

若 $BN_B(x) \neq \Phi$，则称$BN_B(x)$是对象集X关于属性集B的**粗糙集**。

例6.6 对例6.5述信息系统，令X={u1 , u2 ,u3 },求关于条件属性集C的边界区域及其粗集。

解：由上例可得到X关于C的边界区域为：

$BN_C = C^-(X) - C_-(X)$

= {u1 , u2 , u3 , u5} - {u2 , u3}

= {u1 , u5}

由于BNc非空，因此得到对象集X关于属性集C的粗糙集为{u1 , u5}

◆ 回归的含义

回归指研究一组随机变量($Y_1, Y_2, \dots, Y_i$)和另一组($X_1, X_2, \dots, X_k$)变量之间关系的统计分析方法，又称多重回归分析。通常$Y_1, Y_2, \dots, Y_i$是因变量，$X_1, X_2, \dots, X_k$是自变量。

英国统计学家Galton和其学生Pearson研究父母身高与其子女身高的遗传问题，观察了1078对父母，以每对父母身高为X，取他们一个成年儿子的身高为Y，将结果绘成散点图，发现趋势近乎一条直线(父母平均身高68时，成年儿子平均身高69时)：

$$Y = 33.73 + 0.516X$$

可以发现：

$$X = 72;\ Y = 70.89;$$
$$X = 64;\ Y = 66.75$$

6 回归分析的思想

◆ 回归的思想

对变量间统计依赖关系的考察主要是通过相关分析(correlation analysis)或回归分析(regression analysis)来完成的。

6

回归分析的思想

◆ 相关、回归与因果的关系

- 不线性相关并不意味着不相关。
- 有相关关系并不意味着一定有因果关系
- **回归分析/相关分析**研究一个变量对另一个 (些)变量的统计依赖关系，但它们并不意味着一定有因果关系。
- **相关分析**对称地对待任何 (两个)变量，两个变量都被看作是随机的。**回归分析**对变量的处理方法存在不对称性，即区分应变量 (被解释变量)和自变量(解释变量)：前者是随机变量，后者不是。

◆ 回归分析的基本概念

- **回归分析(regression analysis)**是研究一个变量关于另一个(些)变量的具体依赖关系的计算方法和理论。
- 在于通过后者的已知或设定值，去估计和(或)预测前者的(总体)均值。
- 被解释变量(Explained Variable)或应(因)量(Dependent Variable)（前者）。
- 解释变量(Explanatory Variable)或自变量(Independent Variable)（后者）。

回归分析的思想

◆ 回归分析的目的

- 根据样本观察值对经济计量模型参数进行估计，求得回归方程;
- 对回归方程、参数估计值进行显著性检验
- 利用回归方程进行分析、评价及预测。

◆ 总体回归函数(PRF)

■ **回归分析**关心的是根据解释变量的已知或给定值，考察**被解释变量的总体均值**，即当解释变量取某个确定值时，与之统计相关的被解释变量所有可能出现的对应值的平均值。

◆ 总体回归线及函数

- 在给定解释变量X_i条件下被解释变量Y_i的期望轨迹称为**总体回归线** (population regression line)，或更一般地称为**总体回归曲线**(population regression curve)
- 相应的函数:

$$E(Y|X_i)= f(X_i)$$

- 称为(双变量)**总体回归函数** (population regression
- function, **PRF**)。

◆ PRF含义与形式

- **含义**: 回归函数（PRF）说明被解释变量Y的平均状态（总体条件期望）随解释变量X变化的规律。
- **函数形式：**可以是线性或非线性的。
- 示例1中，将每周博彩支出看成是其可支配收入的线性函数时：

$$E(Y|X_i)= B_0 + B_1 X_i$$

为**线性函数**，其中，B_0，B_1是未知参数，称为**回归系数**(regression coefficients)。

6 机器学习

- 机器学习概述
- 记忆学习
- 示例学习
- 决策树学习
- 统计学习
- 集成学习
- 粗糙集知识发现
- 线性回归

  a. 回归分析的思想

  b. 总体回归函数（PRF）的概念

  c. 样本回归函数（SRF）的概念

  d. 普通最小二乘（OLS）方法介绍

◆ 样本回归函数

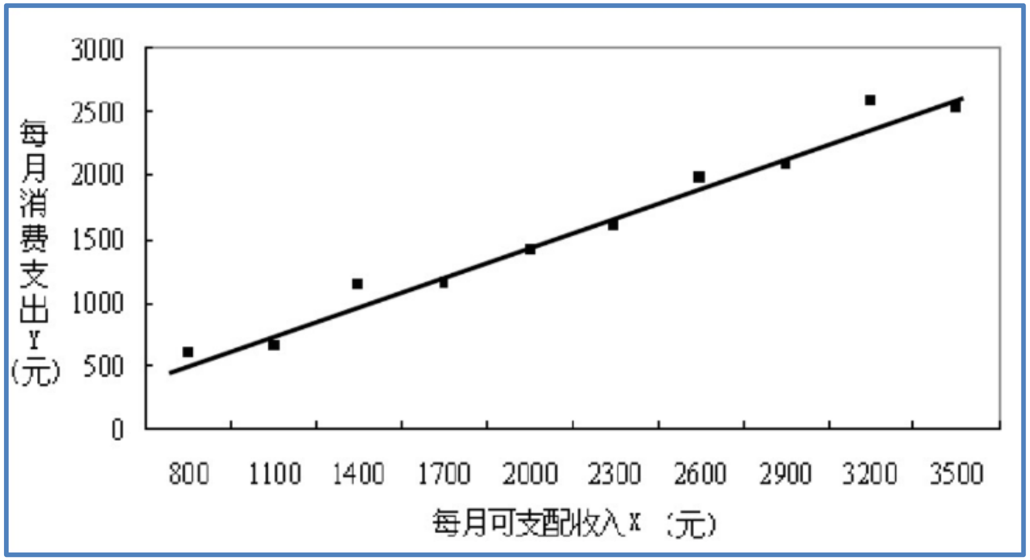
- **问题**：能从一次抽样中获得总体的近似的信息吗?如果可以，如何从抽样中获得总体的近似信息?
- 示例2：假设总体中有如下一个样本，能否从该样本估计**总体回归函数**PRF?

表：家庭消费支出与可支配收入的一个随机样本

Y	800	1100	1400	1700	2000	2300	2600	2900	3200	3500
X	594	638	1122	1155	1408	1595	1969	2078	2585	2530

- 回答：
- -----能

该样本的散点图 (scatter diagram)



画一条直线以尽好地拟合该散点图，由于样本取自总体，可以该直线近似地代表总体回归线。该直线称为样本回归线 (sample regression lines)。

◆ 样本回归函数形式

- 记样本回归线的函数形式为:

$$\widehat{Y}_i = f(X_i) = b_0 + b_1 X_i$$

- 称为**样本回归函数**(sample regression function, SRF)。

- 注意:这里将样本回归线看成总体回归线的近似替代;

$$\widehat{Y}_i = f(X_i) = b_0 + b_1 X_i$$

$$\Downarrow$$

$$Y_i = E(Y|X_i) + u_i = B_0 + B_1 X_i + u_i$$

- 则：$\widehat{Y}_i$是$E(Y|X_i)$的估计量;
- b_i是B_i的估计量， i =0, 1

◆ 样本回归函数的随机形式/样本回归模型

■ 同样地，样本回归函数也有如下的随机形式：

$$\mathrm{Y_i} = \widehat{\mathrm{Y}}_{\mathrm{i}} + \hat{u}_{\mathrm{i}} = b_0 + b_1 \mathrm{X_i} + e_{\mathrm{i}}$$

式中，e_{i}称为样本残差 (或剩余项, Residual)，代表了其他影响$\mathrm{Y_i}$的随机因素的集合可看成是$\hat{u}_{\mathrm{i}}$的估计量u_{i}。

由于方程中引入了随机项，成为计量经济模型，因此也称为样本回归模型 (sample regression model)。

◆ 回归分析的目的

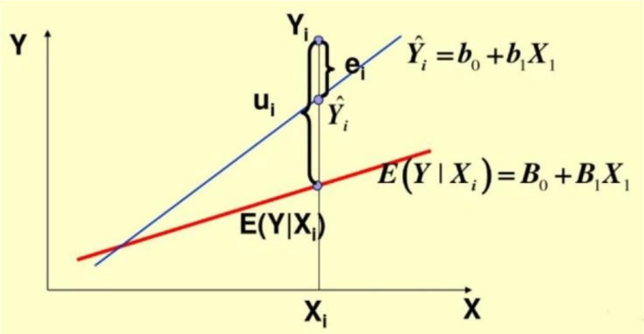
■ 根据样本回归函数SRF，估计总体回归函数PRF：

即，根据 $\mathrm{Y_i} = \widehat{\mathrm{Y}}_\mathrm{i} + e_\mathrm{i} = b_0 + b_1 \mathrm{X_i} + e_\mathrm{i}$

估计 $\mathrm{Y_i}=\mathrm{E(Y|X_i)}+u_\mathrm{i}=B_0 + B_1 \mathrm{X_i}+u_\mathrm{i}$

■ **要求：** 设计适当“方法”构造SRF，以使SRF尽可能“接近”PRF，或者说使b_i(i=0,1)尽可能接近B_i (i=0,1)。

■ **注意：** 这里PRF可能永远无法知道



总体回归线和样本回归线的关系

◆ 参数估计—OLS

说明

- **单方程模型分为两大类：线性模型和非线性模型**

1. **线性模型中，变量之间的关系呈线性关系**
2. **非线性模型中，变量之间的关系呈非线性关系**

- **一元线性回归模型:只有一个解释变量**

$$Y_i = B_0 + B_1 X_i + u_i, i = 1,2,\dots,N$$

Y为被解释变量，X为解释变量，B_0与B_1为待估参数，u为随机干扰项。

◆ 参数估计—OLS

- **回归分析的主要目的**是要通过样本回归函数(模型)SRF尽可能准确地估计总体回归函数(模型) PRF。
- **估计方法**有多种，其中最广泛使用的是**普通最小二乘法**(ordinary least squares, OLS)。
- 为保证参数估计量具有良好的性质，通常对模型提出若干基本假设。
- 实际这些假设与所采用的估计方法紧密相关。

◆ 最小二乘原理

- **对于双变量的PRF**

$$Y_i = B_0 + B_1 X_i + u_i, i = 1,2,\ldots,N$$

- **由于PRF不能直接观察到，故用SRF来估计:**

$$Y_i = b_0 + b_1 X_i + e_i = \widehat{Y}_i + e_i$$

◆ 参数的普通最小二乘估计 (OLS)

- **给定一组样本观测值**$(X_i, Y_i)(i=1,2,\ldots n)$ **要求样本回归函数尽可能好地拟合这组值。**
- **普通最小二乘法 (Ordinary least squares, OLS) 给出的判断标准是：二者之差的平方和**

$$Q = \sum_{i=1}^{n} e^2 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)^2 = [Y_i - (b_0 + b_1 X_i)]^2$$ **最小。**

即在给定样本观测值之下，选择出b_0，b_1，使Y_i与$\hat{Y}_i$之差的 平方和为最小。

◆ 参数的普通最小二乘估计 (OLS)

根据微分运算，可推得用于估计b_0，b_1的下列方程组：

$$\begin{cases}\sum\left(b_0+b_1X_i-Y_i\right)=0\\ \sum\left(b_0+b_1X_i-Y_i\right)X_i=0\end{cases}$$

得：

$$\begin{cases}\sum Y_i=nb_0+b_1\sum X_i\\ \sum X_iY_i=b_0\sum X_i+b_1\sum X_i^2\end{cases}$$

$$\begin{cases}b_0=\dfrac{\sum X_i^2\sum Y_i-\sum X_i\sum X_iY_i}{n\sum X_i^2-\left(\sum X_i\right)^2}=\bar{Y}-b_1\bar{X}\\ b_1=\dfrac{n\sum X_iY_i-\sum X_i\sum Y_i}{n\sum X_i^2-\left(\sum X_i\right)^2}\end{cases}$$

方程组 (*)称为正规方程组 (normal equations).

1 线性回归模型

◆ 参数的普通最小二乘估计 (OLS)

记：

$$\sum x_i^2 = \sum\left(X_i - \bar{X}\right)^2 = \sum X_i^2 - \frac{1}{n}\left(\sum X_i\right)^2$$

$$\sum x_i y_i = \sum\left(X_i - \bar{X}\right)\left(Y_i - \bar{Y}\right) = \sum X_i Y_i - \frac{1}{n}\sum X_i \sum Y_i$$

上述参数估计量可以写成：

$$\begin{cases} b_1 = \dfrac{\sum x_i y_i}{\sum x_i^2} \\ b_0 = \bar{Y} - b_1 \bar{X} \end{cases}$$

称为OLS估计量的离差形式 (deviation form)。

由于参数的估计结果是通过最小二乘法得到的，故称为普通最小二乘估计量 (ordinary leastsquares estimators)，有时也简记为OLS估计量。

1 线性回归模型

◆ OLS参数估计示例

在上述家庭可支配收入-消费支出例中，对于所抽出的一组样本数，参数估计的计算可通过下面的表进行。

参数估计的计算表

	X_i	Y_i	x_i	y_i	$x_i y_i$	x_i^2	y_i^2	X_i^2	Y_i^2
1	800	594	－1350	－973	1314090	1822500	947508	640000	352836
2	1100	638	－1050	－929	975870	1102500	863784	1210000	407044
3	1400	1122	－750	－445	334050	562500	198381	1960000	1258884
4	1700	1155	－450	－412	185580	202500	170074	2890000	1334025
5	2000	1408	－150	－159	23910	22500	25408	4000000	1982464
6	2300	1595	150	28	4140	22500	762	5290000	2544025
7	2600	1969	450	402	180720	202500	161283	6760000	3876961
8	2900	2078	750	511	382950	562500	260712	8410000	4318084
9	3200	2585	1050	1018	1068480	1102500	1035510	10240000	6682225
10	3500	2530	1350	963	1299510	1822500	926599	12250000	6400900
求和	21500	15674			5769300	7425000	4590020	53650000	29157448
平均	2150	1567							

◆ OLS参数估计示例

$$\begin{cases} b_1 = \dfrac{\sum x_i y_i}{\sum x_i^2} = \dfrac{5769300}{7425000} = 0.777 \\ b_0 = \bar{Y} - b_1 \bar{X} = 1567 - 0.777 \times 2150 = -103.172 \end{cases}$$

因此，由该样本估计的回归方程为：

$$\hat{Y}_i = -103.172 + 0.777 X_i$$

课时内容

第7章 支持向量机

7 支持向量机

- 目标：找到一个超平面，使得它能够尽可能多的将两类数据点正确的分开，同时使分开的两类数据点距离分类面最远。

- 解决方法：构造一个在约束条件下的优化问题，具体的说是一个约束二次规划问题(constrained quadratic programing),求解该问题，得到分类器。

7 线性可分与最优分类超平面

◆ 最优分类超平面的定义

假定有以下n个独立、同分布且线性可分的训练样本

$$(x_1, y_1), \quad (x_2, y_2), \quad \ldots, \quad (x_n, y_n)$$

其中，$x_i \in R^n$，n为输入空间的维数；$y_i \in \{-1, +1\}$，表示仅有两类不同的样本。SVM学习的目标就是要找到一个最优超平面

$$w \cdot x + b = 0$$

将两类不同的样本完全分开。其中，w是权重向量，"·"是向量的点积，x是输入向量，b是一个阈值。

如图，其中H为分类超平面，H_1和H_2分别为两个不同类的边界分割平面，H_1和H_2均与H平行，且H_1和H_2分别通过相应类中离H最近的样本点。

对H，若能满足H与两个类边界分割平面H_1和H_2等距，且能使H_1和H_2之间的间隔最大，则称该分类超平面为最优分类超平面。

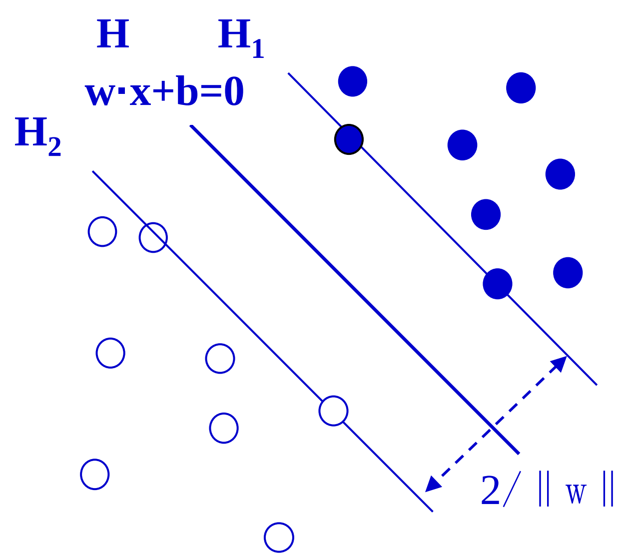


图 线性可分的最优超平面

◆ 最优分类超平面

两点说明：

分类间隔

所谓两个类边界分割平面之间的分类间隔是指它们之间的距离。稍后我们将指出，每个类边界分割平面到最优分类超平面的距离均为

$$1/\|\mathbf{w}\|$$

因此两个类边界分割平面的间隔为

$$2/\|\mathbf{w}\|$$

其中$\|\mathbf{w}\|$是欧氏模函数。

支持向量

最优超平面仅与在H_1和H_2上的训练样本点有关，而与其它训练样本点无关。这些分布在H_1和H_2上的样本点被称为支持向量。因此，

所谓支持向量，就是指那些分布在两个类边界分割平面上的样本点

◆ 最优分类超平面的分类间隔

最优分类超平面作为使分类间隔最大的超平面，可以实现期望风险及结构化风险的最小化。下面对上述类边界分割平面到最优分类超平面的距离，再进一步讨论如下。

对线性分类问题，其分类超平面方程的一般形式为

$$\mathbf{w} \cdot \mathbf{x} + \mathbf{b} = 0$$

由该方程可以得到判别函数的一般形式为

$$\mathbf{g(x)} = \mathbf{w} \cdot \mathbf{x} + \mathbf{b}$$

利用该判别函数，通过对w和b的调整，可将样本空间的样本点分为如下两类：

$$y_i = \begin{cases} +1 & w \bullet x_i + b \geq 0 \\ -1 & w \bullet x_i + b < 0 \end{cases}$$

为使两个不同类中的所有样本都满足|g(x)|≥1，且只有那些离最优分类超平面最近的样本点才有|g(x)|=1，需对判别函数进行归一化处理，使其满足

$$\mathbf{y_i}(\mathbf{w} \cdot \mathbf{x_i} + \mathbf{b}) \geq 1, \qquad \mathbf{i=1,2,\ldots,n}$$

◆ 最优分类超平面的分类间隔

事实上，由于一个样本点到判别式的距离为

$$\frac{|w \bullet x + b|}{\|w\|}$$

且$y_i \in \{-1, +1\}$。因此，一个样本点到判别式的距离可改写为如下形式

$$\frac{y_i(w \bullet x + b)}{\|w\|}$$

我们知道，对线性可分的样本空间，至少应该有一个常D，满足

$$\frac{y_i(w \bullet x + b)}{\|w\|} \geq D$$

我们希望D能够最大化。D越大，说明样本点到分类超平面的距离越大。

最优分类超平面的分类间隔

改变w，可以得到D的无穷多个解。为能够得到其唯一解，我们约定

$$D\|w\| = 1 \quad 或 \quad D = 1/\|w\|$$

此时，要使D最大化，就需要‖w‖最小化。它可归结为如下二次优化问题

$$\min \frac{1}{2}\|w\|^2$$

其约束条件为

$$y_i\,(w \cdot x_i + b) \geq 1, \quad i = 1, 2, \ldots, n$$

通过对上述二次优化问题的求解，即可得到相应的w、b，以及最优分类超平面到类边界分割平面的距离1/‖w‖.

事实上，对n维空间中的线性可分问题，人们已经证明：若输入向量x位于一个半径为r的超球内，则对于满足‖w‖≤A的指示函数集

$$\{ f(x, w, b) = \mathrm{sgn}(w \cdot x + b) \}$$

7 支持向量机

- 线性可分与最优分类超平面
  - a. 最优分类超平面的概念
  - b. 最优分类超平面的分类间隔
  - c. 求解最优分类超平面
- 线性支持向量机
- 非线性支持向量机

7 最优分类超平面的分类间隔

◆ 求解最优分类超平面

求解最优分类超平面，就是要解决前文中给出的二次优化问题，它可通过求解如下拉格朗日函数的鞍点来实现。

$$L(w,b,a)=\frac{1}{2}\|w\|^2-\sum_{i=1}^{n}\alpha_i(y_i(w\bullet x \ +b)-1) \qquad (1)$$

其中α$_i$为拉格朗日乘子，α$_i$≥0，i=1, 2, . . . , n。该二次规划问题存在唯一的最优解。并且，在鞍点上该最优解必须满足对w和b的偏导数为0，即

$$\frac{\partial L}{\partial w}=w-\sum_{i=1}^{n}\alpha_i y_i x_i=0 \qquad (2)$$

$$\frac{\partial L}{\partial b}=\sum_{i=1}^{n}\alpha_i y_i=0 \qquad (3)$$

将(2)、(3)式代入(1)式，消去w、b，即可到原问题的对偶问题：

$$\max W(\alpha)=\sum_{i=1}^{n}\alpha_i-\frac{1}{2}\sum_{i,j=1}^{n}\alpha_i\alpha_j y_i y_j(x_i\bullet x_j) \qquad (7.1)$$

并满足约束条件

$$\sum_{i=1}^{n}\alpha_i y_i=0,\alpha_i\geq 0,i=1,2,\ldots,n$$

7 最优分类超平面的分类间隔

◆ 求解最优分类超平面

可以看出：

使$\alpha_i=0$的样本点对maxW(α)函数没有影响，即对分类不起作用。

仅有使$\alpha_i>0$的样本点才会对分类问题起作用，而这些样本点正是我们所定义的支持向量。

从支持向量开始求最优超平面的主要过程为：

(1) 先从支持向量的样本点中取出任一x_i，求出参数b：

$$b = y_i - w \cdot x_i$$

通常，为了保证稳定性，可对所有支持向量按上式计算，并以其平均值作为b的值。

(2) 然后求出分类判别函数

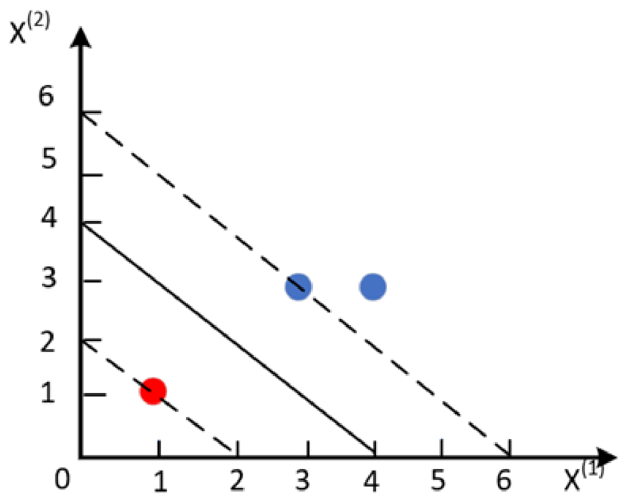
$$f(x) = \operatorname{sgn}(w \bullet x + b) = \operatorname{sgn}\left(\sum_{i=1}^{m} \alpha_i y_i (x_i \bullet x) + b\right) \qquad (7.2)$$

其中，m为支持向量的个数。

这就是线性可分问题的支持向量机。由于有支持向量机所实现的分类超平面具有最大的分类间隔，故相应算法也称为最大间隔算法。

◆ 求解最优分类超平面

例7.1 如图所示的训练数据集，正例点为 $x_1=(3,3)^T$，$x_2=(4,3)^T$，负例点为 $x_3=(1,1)^T$，求最大间隔分离超平面。



解法1：约束二次最优化问题：

$$\min_{\omega,b}\frac{1}{2}(\omega_1^2+\omega_2^2)$$

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

求得该最优化问题的解为 $\omega_1=\omega_2=\frac{1}{2}$，$b=-2$，所以最大间隔分离超平面为 $\frac{1}{2}x^{(1)}+\frac{1}{2}x^{(2)}-2=0$，$x_1$ 和 x_3 为支持向量。

◆ 求解最优分类超平面

解法2：对偶问题

$$\min_{\alpha} \frac{1}{2}\sum_{i,j=1}^{3}\alpha_i\alpha_j y_i y_j (x_i \cdot x_j) - \sum_{i=1}^{3}\alpha_i$$

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

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

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

把 $\alpha_3 = \alpha_1 + \alpha_2$ **代入目标函数并记为**

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

$s(\alpha_1, \alpha_2)$ **对** α_1 **和** α_2 **求偏导数并令其等于零得**

$$\frac{\partial s}{\partial \alpha_1} = 8\alpha_1 + 10\alpha_2 - 2 = 0$$

◆ 求解最优分类超平面

$$\frac{\partial s}{\partial \alpha_2} = 13\alpha_2 + 10\alpha_1 - 2 = 0$$

解得$s(\alpha_1,\alpha_2)$在$(\frac{3}{2},-1)^T$处取得极值，但不满足$\alpha_2 \geq 0$，所以最小值应在边界上取到。

因为$s(0,\frac{2}{13}) = -\frac{2}{13}$，$s(\frac{1}{4},0) = -\frac{1}{4}$，因此$s(\alpha_1,\alpha_2)$在$(\frac{1}{4},0)^T$处取得极小值，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和x_3是支持向量。计算得

$$\begin{pmatrix} \omega_1^* \\ \omega_2^* \end{pmatrix} = \frac{1}{4}\begin{pmatrix} 3 \\ 3 \end{pmatrix} + 0 - \frac{1}{4}\begin{pmatrix} 1 \\ 1 \end{pmatrix} = \begin{pmatrix} \frac{1}{2} \\ \frac{1}{2} \end{pmatrix}$$

◆ 求解最优分类超平面

若取 $j=1$，（对应X_3），则

$$b^* = 1 - \frac{1}{2} * 3 - \frac{1}{2} * 3 = -2$$

若取 $j=3$（对应X_1），则

$$b^* = -1 - \frac{1}{2} * 1 - \frac{1}{2} * 1 = -2$$

最终求得分离超平面：

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

分类决策函数：

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

可以看到所得结果与解法 1 完全一致。