课程简介

数学与计算机交叉的基础课程，主要介绍如何将现实中的决策问题求解抽象为**数学优化模型**，并通过系统的方法求得最优解。

课程既注重**理论基础**，又强调**方法应用**，旨在为后续机器学习、深度学习及计算机视觉等课程奠定坚实的理论与方法基础。

① *掌握优化问题的建模思路与基本理论;*

② *理解常见优化算法的基本原理、适用范围;*

③ *能够将数学优化方法应用到工程、数据分析等实际问题中;*

④ *为后续深度学习类课程提供数学与方法论支撑。*

第一章 最优化简介

目录

§1.1 最优化问题概括

§1.1 最优化问题概括

最优化问题泛指定量决策问题，主要关心如何对**有限资源**进行有效分配和控制，并达到某种意义上的最优。

- **最优化问题一般可以描述为：**

$$\min f(x)$$
$$s.t. \quad x \in \mathbb{X}$$

① $x = (x_1, x_2, \ldots, x_n)^T \in \mathbb{R}^n$ 是**决策变量**

② $f : \mathbb{R}^n \longrightarrow \mathbb{R}$ 是**目标函数**

③ $\mathbb{X} \subseteq \mathbb{R}^n$ 是**约束集合或可行域**

④ 可行域中的点称为**可行解或可行点**

§1.1 最优化问题概括

- 求解 $\min f(x_1, x_2) = x_1^2 + x_2^2 - 2x_1 - 4x_2$

$$s.t. \quad -x_1 - x_2 + 1 \geq 0, \quad x_1 \geq 0, \; x_2 \geq 0$$

目标函数

决策变量

可行域、可行解

§1.2 最优化的基本概念

§1.2 最优化问题概括

- 无约束和约束最优化问题
- 线性和非线性规划问题
- 凸和非凸优化问题
- 连续和离散优化问题
- 随机和确定性优化问题
- 全局和局部最优解

无约束和约束最优化问题

无约束优化问题泛指目标函数在**整个实数空间**中寻找最优解，不存在任何约束条件。

- **定义**：

$$\min f(x)$$
$$s.t. \quad x \in \mathbb{R}^n$$

- 可行域：整个空间 $\mathbb{R}^n$
- 常用方法：梯度下降法、牛顿法、拟牛顿法
- 示例：最小二乘回归

无约束和约束最优化问题

无约束优化问题

- **问题**：

$$\min f(x) = (x-2)^2$$
$$s.t. \quad x \in \mathbb{R}$$

- **解析**：
  - 没有约束条件；
  - 最优解： $x = 2$，最小值为0。
- 函数图像是一条抛物线，最小点在$x = 2$，对应的最小值为0

习题1

无约束和约束最优化问题

约束优化问题泛指在给定约束条件下最小化（或最大化）目标函数。

- **定义**：

$$\min f(x)$$

$$s.t.\quad g_i(x) \geq 0, i \in I$$
$$\qquad h_j(x) = 0, j \in E$$

- **可行域**：由约束函数决定

$$\mathbb{X} = \{x \in \mathbb{R}^n | g_i(x) \geq 0, i \in I; h_j(x) = 0, j \in E\}$$

无约束和约束最优化问题

等式约束优化： $E \neq \emptyset, \quad I = \emptyset$

- **定义：**

$$\min f(x)$$
$$s.t. \quad h_j(x) = 0$$

- **可行域：**

$$\mathcal{X} = \{x | h_j(x) = 0, j \in E\}$$

- **求解方法：** 拉格朗日乘子法

- **拉格朗日函数：**

$$L(x, \lambda) = f(x) + \sum \lambda_j h_j(x)$$

无约束和约束最优化问题

等式约束优化：以构造拉格朗日乘子法为例

- **问题：**

$$\min f(x, y) = x^2 + y^2$$
$$s.t. \quad x + y = 1$$

- **构造拉格朗日函数：**

$$L(x, y, \lambda) = x^2 + y^2 + \lambda(x + y - 1)$$

解方程组可得

$$x = y = \frac{1}{2}, \quad f(x, y) = \frac{1}{2}$$

- **几何意义：** 在直线$x + y = 1$上寻找距离原点最近的点

习题2

无约束和约束最优化问题

不等式约束优化： $E = \emptyset, \ I \neq \emptyset$

- **定义：**

$$\min f(x)$$
$$s.t. \quad g_i(x) \geq 0$$

- **可行域：**

$$\mathcal{X} = \{x | g_i(x) \geq 0, i \in I\}$$

- **求解方法：** KKT条件、内点法

无约束和约束最优化问题

不等式约束优化：以KKT条件为例

- **问题：**

$$\min f(x) = x^2$$
$$s.t. \quad x \geq 1$$

- **解析：**
  - 无约束最优解在$x = 0$，但不满足约束条件；
  - 在可行域$x \geq 1$上，最优解为$x = 1$，函数值为$f(x) = 1$。
- **几何意义：** 抛物线在$x = 1$处截断，最小点被“推”到边界。

习题3

无约束和约束最优化问题

一般约束优化: $E \neq \emptyset, \quad I \neq \emptyset$

- **定义**: 同时存在等式和不等式约束，即:

$$\min f(x)$$
$$s.t. \quad g_i(x) \geq 0, i \in I; \mathrm{h}_j(x) = 0, j \in E$$

- **可行域**:

$$\mathbb{X} = \{x \in \mathbb{R}^n | g_i(x) \geq 0, i \in I; \mathrm{h}_j(x) = 0, j \in E\}$$

- **求解方法**: 罚函数法、增广拉格朗日法、内点法

无约束和约束最优化问题

一般约束优化：以罚函数法为例

- **问题**：

$$\min f(x,y) = x^2 + y^2$$
$$s.t.\quad x + y = 1, x \geq 0$$

- **罚函数法转化**：

$$F(x,y) = f(x,y) + \mu(x+y-1)^2 + \mu\max(0,-x)^2$$

其中μ为惩罚系数

- **解析**：
  - 当μ很大时，解会趋近于满足约束；
  - 数值优化可得到近似解$x = 0.5, y = 0.5$

习题4

- **思想**：违反约束就“罚”，引导解逐步进入可行域

线性和非线性规划问题

线性规划指目标函数与约束条件函数**均为线性函数**的优化问题。

- **定义**：

$$\min \ c^T x$$
$$s.t. \quad Ax = b, x \geq 0$$

其中： $A \in \mathbb{R}^{m\times n}, b \in \mathbb{R}^m$， c^T**表示**$c \in \mathbb{R}^n$**的转置。**

- **目标函数**： $c^T x$
- **约束条件**：$Ax = b$（线性等式约束）
- **非负约束**： $x \geq 0$
- **求解方法**： 单纯形法、内点法

线性和非线性规划问题

线性规划指目标函数与约束条件函数**均为线性函数**的优化问题。

- **问题：**

$$\min 3x_1 + 5x_2$$

$$s.t.\quad \begin{cases} 2x_1 + x_2 \le 100 \\ x_1 + 3x_2 \le 90 \\ x_1, x_2 \ge 0 \end{cases}$$

- **候选顶点：**(0,0), (50,0), (0,30), (42,16)
- **对应目标函数值：** 0, 150, 150, 206
- **最优解：** $x^* = (0,0), f^* = 0$； *表示最优解

习题5

线性和非线性规划问题

非线性规划指**目标函数或约束条件函数**中至少有一个是**非线性函数**的优化问题。

- **定义**：

$$\min \ f(x)$$
$$s.t. \ \ g_i(x) \geq 0, i \in I; \mathrm{h}_j(x) = 0, j \in E$$

- **求解方法**：梯度法、牛顿法、内点法、序列二次规划

- **示例**：

$$\min \ f(x, y) = x^2 + y^2$$
$$s.t. x^2 + y \leq 1, x \geq 0$$

可行域是曲线和区域的交集，解通常在边界。

习题6

线性和非线性规划问题

二次规划指**目标函数为二次函数**，**且约束条件为线性函数**的优化问题；二次规划是一种特殊的非线性规划。

- **定义**：

$$\begin{aligned} \min \quad & \frac{1}{2}x^TQx + c^Tx \\ s.t. \quad & Ax = b, x \geq 0 \end{aligned}$$

其中：$Q \in \mathbb{R}^{n\times n}$,$A \in \mathbb{R}^{m\times n}$, 且$b \in \mathbb{R}^m$。

- **示例**：支持向量机

$$\begin{aligned} \min \quad & \frac{1}{2}\|w\|^2 \\ s.t. \quad & y_i(w^Tx_i + b) \geq 1 \end{aligned}$$

凸和非凸优化问题

凸优化问题指目标函数是凸函数，且可行域是凸集的优化问题。

- **定义**：

$$\min f(x)$$
$$s.t.\ \ g_i(x) \geq 0, h_j(x) = 0$$

其中：f 为凸函数，约束形成的集合为凸集

- **特点**：局部最优解即全局最优解
- **典型示例**：
  - 最小二乘回归
  - LASSO回归
  - 支持向量机

凸和非凸优化问题

非凸优化问题指目标函数不是凸函数，或可行域不是凸集的优化问题。

- **特点：**
  - 存在多个局部最优解
  - 理论分析非常困难
  - 求解算法一般只能保证找到“局部最优”或“近似最优”解
- **在现实世界中应用广泛：**
  - 机器学习：聚类（如K-means）算法
  - 深度学习：神经网络训练（非凸损失函数）
  - 组合优化：旅行商问题、背包问题等

连续和离散优化问题

连续优化问题指目标函数和约束条件都是连续函数的优化问题；决策变量可以在一个连续空间(通常是实数域$\mathbb{R}^n$)中取值。

- **特点**：变量取值是连续的，可以用微积分工具（如梯度）求解
- **求解方法**：梯度下降、牛顿法、拟牛顿法、内点法等
- **示例**：最小二乘回归

$$\min \ \frac{1}{2}\|y - Xw\|^2$$

连续和离散优化问题

离散优化问题指目标函数或约束条件中至少有一个是离散的；决策变量的取值范围是离散值（如整数、有限集合）。

- **特点**：变量取值是离散的，通常不能直接用微积分工具
- **求解方法**：多数为NP-hard问题，常用近似算法或启发式方法
- **示例**：
  - **旅行商问题**：寻找最短路径经过所有城市
  - **背包问题**：有限物品，选择是否放入背包以最大化价值

随机和确定性优化问题

随机优化指在目标函数或约束条件中包含随机性或不确定性时，采用概率方法或随机算法来寻找最优解。

$$\min_{\omega} F(\omega) = \mathbb{E}_{\xi \sim \mathcal{D}}[f(\omega; \xi)]$$

- ω：待优化参数
- ξ：随机变量（比如训练样本、噪声）
- $\mathcal{D}$：数据或者噪声的分布
- $f(\omega; \xi)$：在单一样本或环境下的损失

随机和确定性优化问题

确定性优化指在目标函数和约束条件已知且确定的情况下，使用解析方法或数值方法寻找最优解。

- **定义**：

$$\min f(x)$$
$$s.t.\ \ g_i(x) \le 0,\ i \in I, \mathrm{h}_j(x) = 0, j \in E$$

其中： $f(x)$， $g_i(x)$, $\mathrm{h}_j(x)$ 都是确定的函数

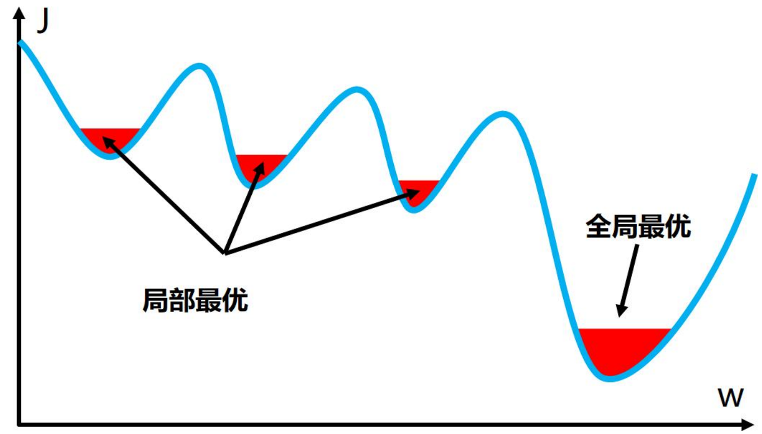
全局和局部最优解

- **定义：**

点x^*称为**局部最优解**，如果在某个领域$N(x^*)$内：

$$f(x^*) \leq f(x), \forall x \in N(x^*)$$

- 局部最优解只在**小范围**内最优
- 非凸优化问题常有多个局部最优解
- 在实际优化中，梯度法往往只能保证收敛到局部最优



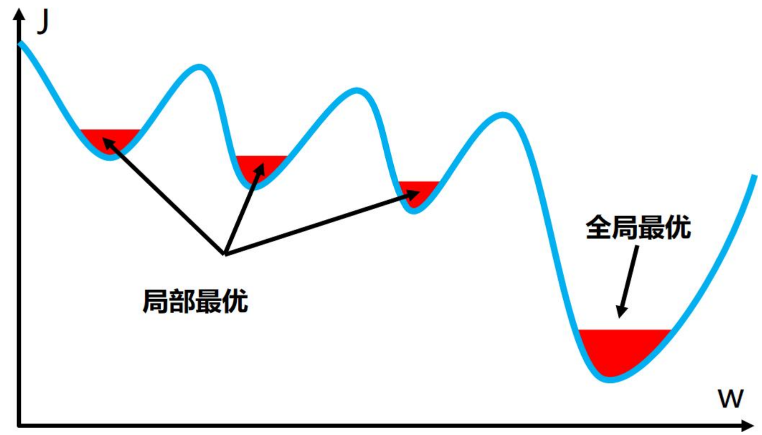
全局和局部最优解

- **定义**:

点x^*称为全局**最优解**，如果在整个可行域F内:

$$f(x^*) \le f(x), \forall x \in F$$

- 全局最优解是整个问题的最优解
- 凸优化问题保证局部最优解=全局最优解
- 非凸优化问题中，找到全局最优很困难



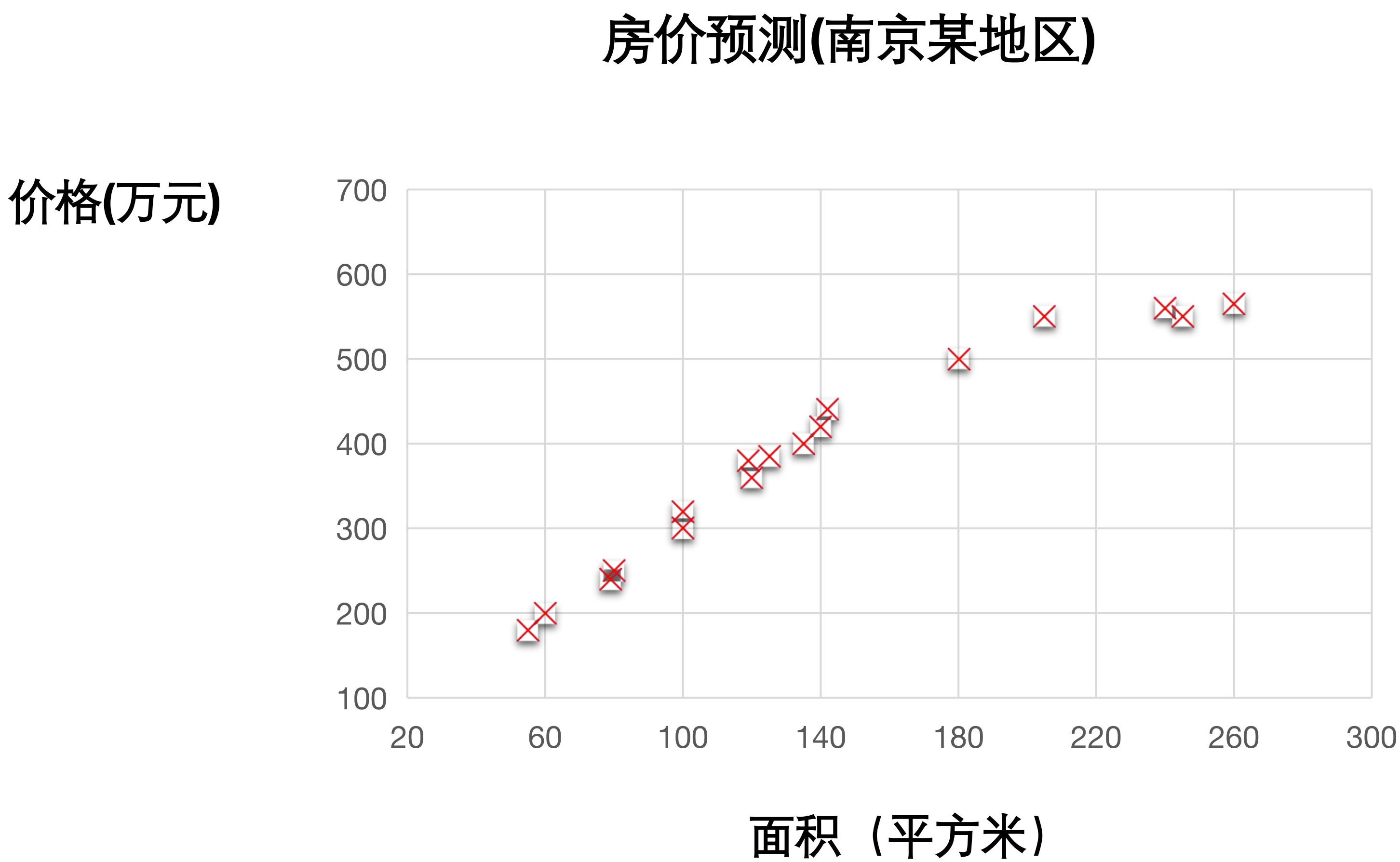
§1.3 优化实例

§1.3 优化实例

- 最小二乘线性回归
- LASSO（套索回归）
- 支持向量机
- 多层感知机

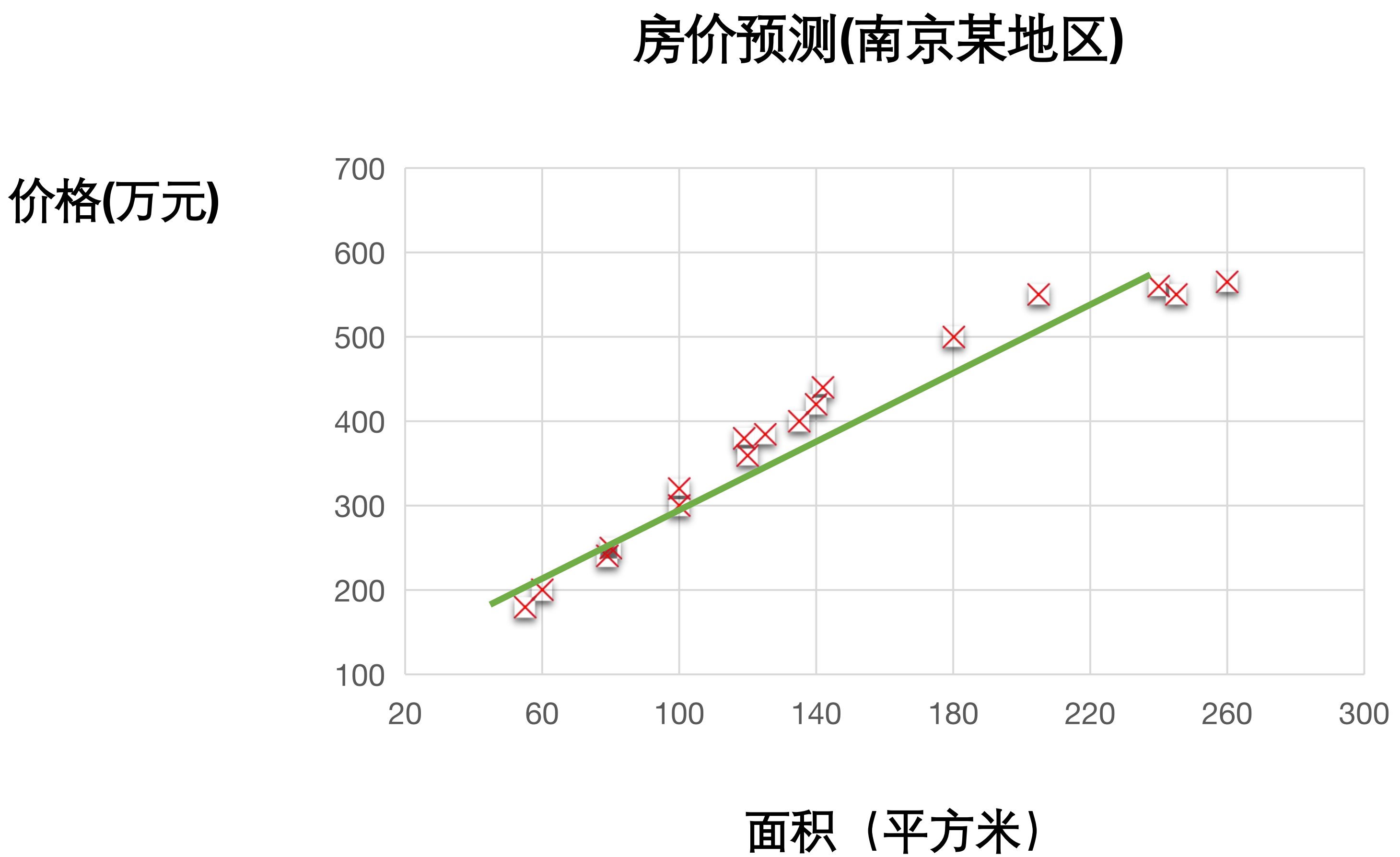
最小二乘线性回归

- 示例(回归问题)



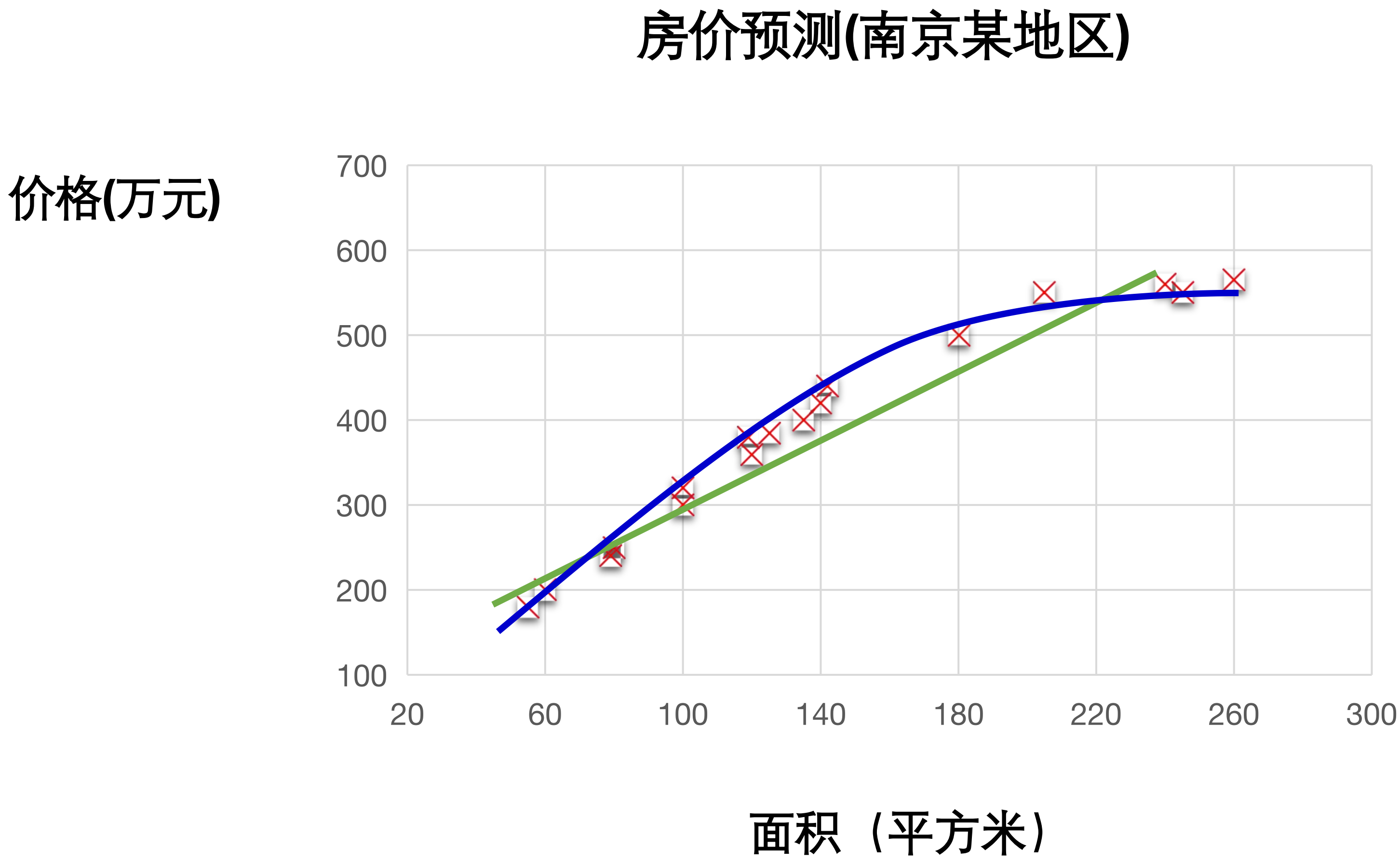
最小二乘线性回归

- 示例(回归问题)



最小二乘线性回归

- 示例(回归问题)



最小二乘线性回归

- 示例(回归问题)

- 训练集：
  - $\boldsymbol{X}_i \in \mathbb{R}^n$：输入变量（特征）
  - $\boldsymbol{y}_i \in \mathbb{R}$：输出变量（目标值）
  - $(\boldsymbol{X}, \boldsymbol{y})$代表训练集中的实例或样本；
  - $(\boldsymbol{X}_i, \boldsymbol{y}_i)$代表第 i 个观察实例或样本；
  - $i = 1, 2, \ldots, m$代表训练集中样本的数量。

面积（平方米）	价格（万元）
60	200
80	240
100	300
100	306
…	...

最小二乘线性回归

线性回归研究输入变量和输出变量之间的线性关系。在给定训练集下，拟合一个**线性函数**，通过最小化**误差的平方和**，使得预测值与真实值之间的误差最小。

- **训练数据：** m个训练样本点

$$\{(X_i, y_i)_{i=1}^m\} \quad , \quad X_i \in \mathbb{R}^n, \quad y_i \in \mathbb{R}$$

- **数据矩阵表示**

$$X = \begin{bmatrix} x_1^T \\ x_2^T \\ \cdots \\ x_m^T \end{bmatrix} \in \mathbb{R}^{m \times n} \qquad y = \begin{bmatrix} y_1 \\ y_2 \\ \cdots \\ y_m \end{bmatrix} \in \mathbb{R}^m$$

- **超平面（目标函数）**

$$\mathrm{h}(z) = w^T z + b$$

其中： $w \in \mathbb{R}^n$ 表示权重向量； $b \in \mathbb{R}$ 表示偏置项

最小二乘线性回归

线性回归研究输入变量和输出变量之间的线性关系。在给定训练集下，拟合一个线性函数，通过最小化误差的平方和，使得预测值与真实值之间的误差最小。

- **引入增广向量**
  - **增广数据点：** $\tilde{\mathbf{z}} := (1, \mathbf{z}^T)^T \in \mathbb{R}^{n+1}$
  - **增广权重向量：** $\widetilde{\mathbf{w}} := (\mathbf{b}, \mathbf{w}^T)^T \in \mathbb{R}^{n+1}$
- **目标函数**

$$\mathrm{h}(\mathrm{z}) = w^T z + b \quad \Longrightarrow \quad \mathrm{h}(\mathrm{z}) = \widetilde{w}^T \tilde{z}$$

- **预测向量**

$$\hat{y} = \tilde{D}\widetilde{w}$$

- **优化目标**

$$min\|y - \tilde{D}\widetilde{w}\|$$

最小二乘线性回归

小结

- 根据跟定的数据点，寻找最优超平面或增广权重向量来最小化预测值$\hat{y}$和真实值y之间的误差。
- 优化目标：

$$\min \|\hat{y}-y\|_2^2 = \min \|\tilde{D}\tilde{w}-y\|_2^2$$

- 最优解（正规方程）：

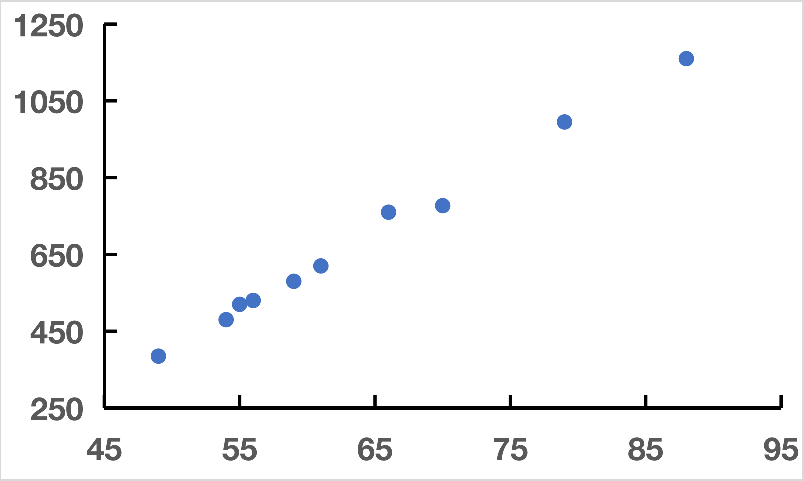
$$\tilde{w} = (\tilde{D}^T\tilde{D})^{-1}\tilde{D}y$$

最小二乘线性回归

练习题

某小区房价与面积样本数据如下表所示，请选择一些数据样本（如序号1-5）建立房价与面积的**线性回归模型**，并计算其均方误差MSE。

某小区房价		
序号	价格/万	面积
1	385	49
2	480	54
3	520	55
4	530	56
5	580	59
6	620	61
7	760	66
8	777	70
9	995	79
10	1160	88



最小二乘线性回归

练习题

某小区房价与面积样本数据如下表所示，请选择一些数据样本（如序号1-5）建立房价与面积的**线性回归预测模型**，并计算其均方误差MSE。

某小区房价		
序号	价格/万	面积
1	385	49
2	480	54
3	520	55
4	530	56
5	580	59

解：

利用前5组数据，基于最小二乘法，线性模型的系数为：

$$\widetilde{w} = (\tilde{D}^T \tilde{D})^{-1} \tilde{D} y$$

$$D = \begin{bmatrix} 49 & 1 \\ 54 & 1 \\ 55 & 1 \\ 56 & 1 \\ 59 & 1 \end{bmatrix} \quad y = \begin{bmatrix} 385 \\ 480 \\ 520 \\ 530 \\ 580 \end{bmatrix}$$

由此计算出：

$w = 19.89, b =- 586.84$,MSE=7.5

线性回归模型为：

$$y = 19.89x - 586.84$$

习题7

§1.3 优化实例

- 最小二乘线性回归
- LASSO
- 支持向量机
- 多层感知机

LASSO(索套回归)

- **定义：**
  - LASSO（Least Absolute Shrinkage and Selection Operator）是一种线性回归方法，又称为L1正则化。
  - 在最小二乘目标中加入L1正则化，能够：1）限制参数大小，防止过拟合；2）产生稀疏解，即部分权重收缩为0，实现特征选择。
- **数学表达：**

$$\min \|\tilde{D}\tilde{w} - y\|_2^2 + \lambda \|\tilde{w}\|_1$$

  - 第一项：拟合误差（残差平方和）
  - 第二项：L1正则化项（参数的绝对值和）
  - $\lambda > 0$：正则化系数
- **与L2正则化对比：**
  - L2：收缩参数，但很少为0
  - L1：推动部分参数为0，得到稀疏解

LASSO(索套回归)

- **求解方法:**
  - 梯度下降 (Gradient Descent)
  - 软阈值迭代 (Soft-thresholding / ISTA)
  - 神经网络近似方法(如深度展开算法, Learned ISTA)

- **以软阈值迭代为例,对于每一步迭代:**

$$w^{k+1} = S_{\frac{\lambda}{2}}(w^k - \frac{1}{L}D^T(Dw^k - y))$$

  - L: 步长参数
  - $S_\theta(*)$: 软阈值算子 (soft-thresholding operator)

$$S_\theta(z) = sign(z) \cdot \max(|z| - \theta, 0)$$

§1.3 优化实例

- 最小二乘线性回归
- LASSO
- 支持向量机
- 多层感知机

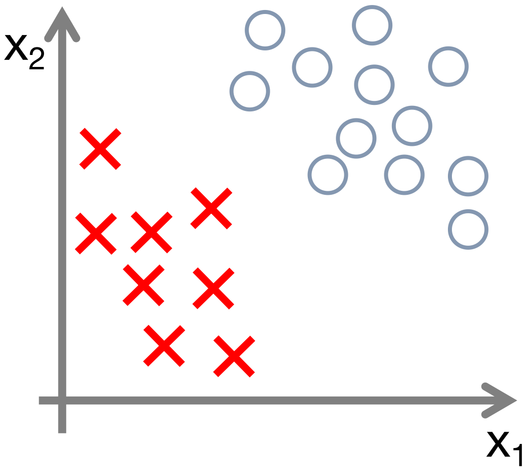
支持向量机（Support Vector Machine, SVM）

- **定义：**
  - 支持向量机（Support Vector Machine, SVM）是一类用于分类和回归的监督学习模型
  - 核心思想：找到一个最优超平面，使两类数据的间隔最大

- **分类：**
  - **硬间隔SVM**
    - 适用于线性可分的情况
    - 要求所有样本完全正确分类
  - **软间隔SVM**
    - 允许一定的分类错误
    - 通过引入松弛变量，解决线性不可分问题
  - **核SVM**
    - 利用核函数将数据映射到高维特征空间
    - 适用于非线性分类

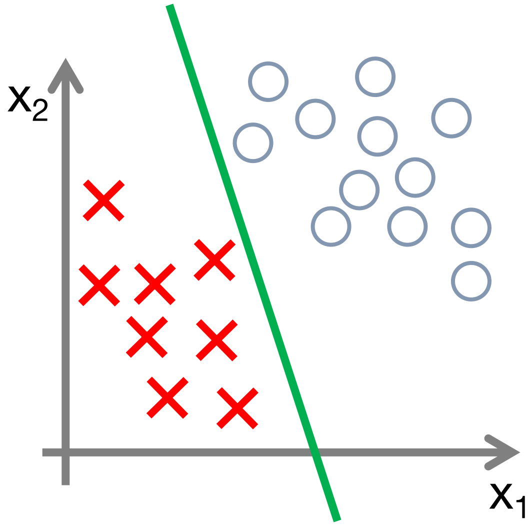
硬间隔支持向量机

例：一组叉和圆的样本数据如下所示，请确定一个超平面（或分界线）将叉与圆这两类样本分隔开，并思考这样的分割面/线有多少个？什么样的分割面/线是最佳的？为什么？



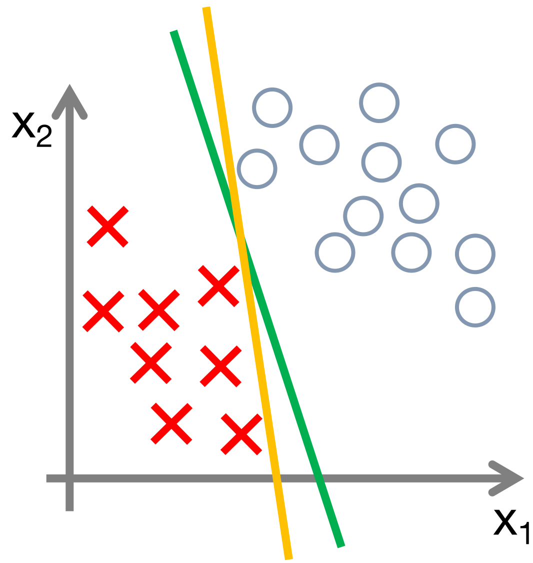
硬间隔支持向量机

例：一组叉和圆的样本数据如下所示，请确定一个超平面（或分界线）将叉与圆这两类样本分隔开，并思考这样的分割面/线有多少个？什么样的分割面/线是最佳的？为什么？



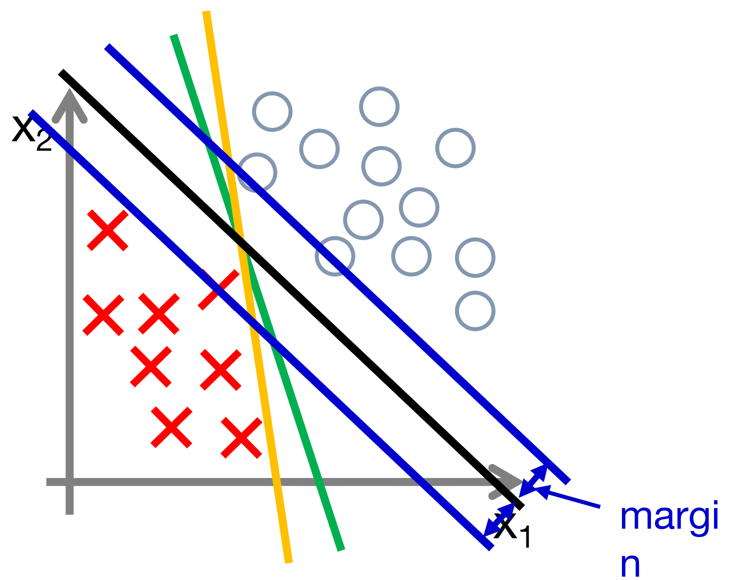
硬间隔支持向量机

例：一组叉和圆的样本数据如下所示，请确定一个超平面（或分界线）将叉与圆这两类样本分隔开，并思考这样的分割面/线有多少个？什么样的分割面/线是最佳的？为什么？



硬间隔支持向量机

例：一组叉和圆的样本数据如下所示，请确定一个超平面（或分界线）将叉与圆这两类样本分隔开，并思考这样的分割面/线有多少个？什么样的分割面/线是最佳的？为什么？



最大化最近的数据点(或类)和超平面之间的距离，将帮助我们确定最佳的分类超平面

SVM：最大间距分类器（Large margin classifier）

硬间隔支持向量机

线性可分二分类硬间隔支持向量机

假设具有正负标签的二分类数据是线性可分的。

- **超平面：**

$$\mathrm{h}(x) = \omega^T x + b$$

$$\Rightarrow \quad y = \mathrm{sgn}\,(\mathrm{h}(x)) = \begin{cases} +1, & \mathrm{h}(x) > 0, \\ -1, & \mathrm{h}(x) < 0 \end{cases}$$

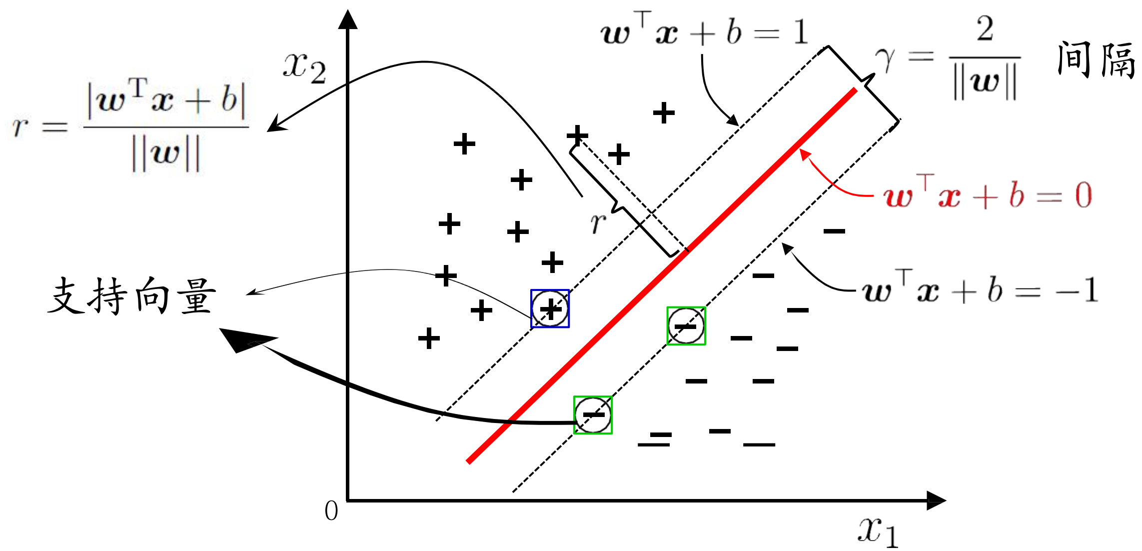
- **分类条件：**

$$y_i(\omega^T x_i + b) \geq 1, i = 1,2,\ldots,m$$

硬间隔支持向量机

线性可分二分类硬间隔支持向量机

超平面方程：$h(x) = \omega^T x + b = 0$ **x到超平面的距离**：$r = \frac{|\boldsymbol{w}^{\mathrm{T}}\boldsymbol{x} + b|}{||\boldsymbol{w}||}$



- **支持向量**：距离超平面最近的几个训练样本；
- **间隔**：两个不同类支持向量到超平面的距离之和 $r = \frac{2}{\|\omega\|}$

硬间隔支持向量机

线性可分二分类硬间隔支持向量机

- **优化目标**：寻找参数ω和b，以最大化间隔$r = \frac{2}{\|\omega\|}$。

$$\max_{\omega,b} \frac{2}{\|\omega\|}$$

$$s.t. \quad y_i(\omega^T x_i + b) \geq 1, i = 1,2,\ldots,m$$

等价于

$$\min_{\omega,b} \frac{1}{2}\|\omega\|^2$$

$$s.t. \quad y_i(\omega^T x_i + b) \geq 1, i = 1,2,\ldots,m$$

- 这就是支持向量机的基本型；
- 凸二次规划问题，能用优化计算包求解，但可以有更高效的办法。

硬间隔支持向量机

线性可分二分类硬间隔支持向量机

- **拉格朗日乘子法**

  - **第一步**：引入拉格朗日乘子a_i，得到拉格朗日函数：

$$L(\omega, b, a) = \frac{1}{2}\|\omega\|^2 - \sum_{i=1}^{m} a_i(y_i(\omega^T x_i + b) - 1)$$

  - **第二步**：令$L(\omega, b, a)$对ω和b的偏导为零，可得：

$$\boldsymbol{w} = \sum_{i=1}^{m} \alpha_i y_i \boldsymbol{x}_i \ , \qquad 0 = \sum_{i=1}^{m} \alpha_i y_i$$

  - **第三步**：将第二步中的公式回代入第一步，可得：

(**对偶问题**)

$$\max_{\boldsymbol{\alpha}} \ \sum_{i=1}^{m} \alpha_i - \frac{1}{2} \sum_{i=1}^{m} \sum_{j=1}^{m} \alpha_i \alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}} \boldsymbol{x}_j$$

$$\text{s.t.} \ \sum_{i=1}^{m} \alpha_i y_i = 0 \ , \ \alpha_i \geqslant 0 \ , \quad i = 1, 2, \ldots, m$$

硬间隔支持向量机

线性可分二分类硬间隔支持向量机

- 先最小化（找到最优超平面），再最大化（找到约束最紧的点，即支持向量）

$$\max_{\alpha}\min_{w,b} L(w,b,\alpha)$$

- KKT条件：最优解的必要条件（在凸问题中也是充分条件）

$$\begin{cases} a_i \geq 0 \\ y_i(\omega^T x_i + b) - 1 \geq 0 \\ a_i(y_i(\omega^T x_i + b) - 1) = 0 \end{cases} \Rightarrow \begin{array}{l} \text{必有 } a_i = 0, \\ \text{或 } y_i(\omega^T x_i + b) = 1 \end{array}$$

解的稀疏性：训练完成后，最终模型仅与支持向量有关

支持向量机（SVM）因此而得名。

硬间隔支持向量机

例：给出标签为+1的的两个点$x_1=(0,0)^T$和$x_2=(1,1)^T$，以及标签为-1的一个点$x_3=(0,2)^T$，求硬间隔线性支持向量机。

解：记x_1，x_2，x_3对应的标签分别为y_1，y_2，y_3，得$y_1=y_2=1$，$y_3=-1$

代入公式：

$$\max_{\boldsymbol{\alpha}} \sum_{i=1}^{m}\alpha_i-\frac{1}{2}\sum_{i=1}^{m}\sum_{j=1}^{m}\alpha_i\alpha_j y_i y_j \boldsymbol{x}_i^{\mathrm{T}}\boldsymbol{x}_j$$

得：

$$\begin{aligned} L(a_1,a_2,a_3) &= a_1+a_2+a_3-\frac{1}{2}(a_2^2y_2^2x_2^Tx_2+2a_2a_3y_2y_3x_2^Tx_3+a_3^2y_3^2x_3^Tx_3) \\ &= a_1+a_2+a_3-\frac{1}{2}(2a_2^2-4a_2a_3+4a_3^2) \\ &= a_1+a_2+a_3-\frac{1}{2}(2a_2^2-4a_2a_3+4a_3^2) \end{aligned}$$

约束条件：

$$\sum_{i=1}^{3}a_iy_i=a_1+a_2-a_3=0 \Rightarrow a_1+a_2=a_3$$

代入上式：

$$L(a_1,a_2)=2a_1+2a_2-2a_1^2-a_2^2-2a_1a_2$$

习题8

硬间隔支持向量机

例：给出标签为+1的的两个点$x_1 = (0,0)^T$和$x_2 = (1,1)^T$，以及标签为-1的一个点$x_3 = (0,2)^T$，求硬间隔线性支持向量机。

解：
$$L(a_1, a_2) = 2a_1 + 2a_2 - 2a_1^2 - a_2^2 - 2a_1a_2$$

求偏导，并令其为0：

$$\begin{cases} \dfrac{\partial L}{\partial a_1} = 2 - 4a_1 - 2a_2 = 0 \\ \dfrac{\partial L}{\partial a_2} = 2 - 2a_2 - 2a_1 = 0 \end{cases} \Rightarrow \begin{matrix} a_1 = 0 \\ a_2 = 1 \end{matrix} \Rightarrow a_3 = 1$$

分离超平面和偏置分别为：

$$\omega = \sum_{i=1}^{3} a_i y_i x_i = (1, -1)^T, \quad b = \frac{(y_2 - \omega^T x_2) + (y_3 - \omega^T x_3)}{2} = 1$$