
# Python
# 机器学习实战

# 线性回归和梯度下降

# 线性回归

## 预测

### 连续型变量

回归（线性回归，广义线性回归、CART）

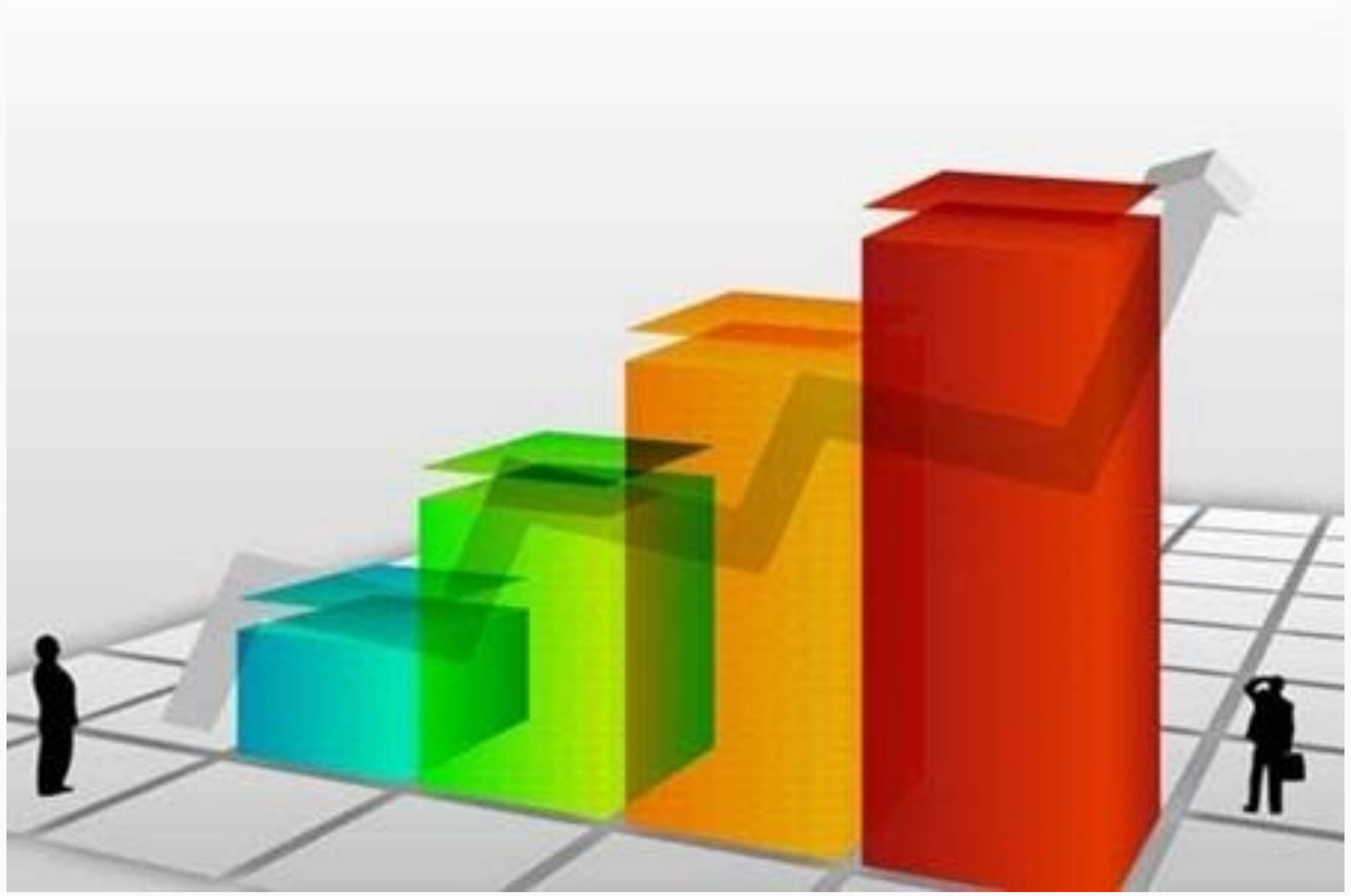
### 离散型变量

分类（逻辑回归、SVM、决策树…）

## 广告费和销量

- 销量随着广告费的投入增加而增加吗?
- 能不能定量的描述广告投入和销量之间的关系?
- 广告费对于销量的影响有统计意义吗?
- 贡献有多大?
- 能不能根据广告投入预测销量?

… …



# 线性回归

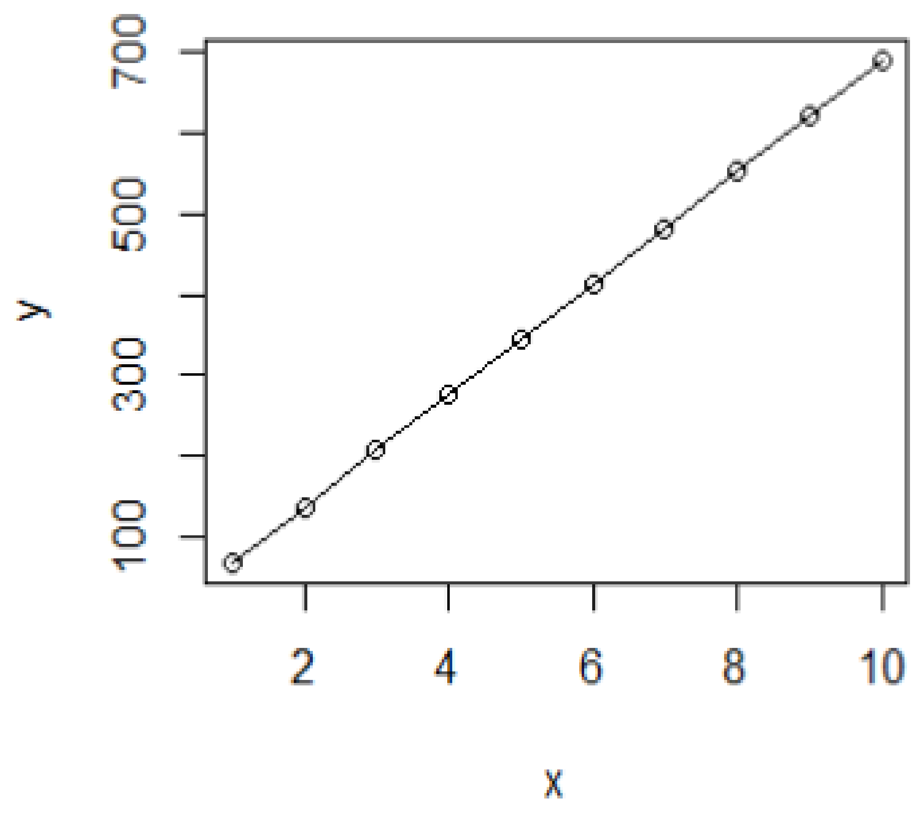
不独立两个随机变量，二者之间肯定会存在某种关系：

## 确定性关系

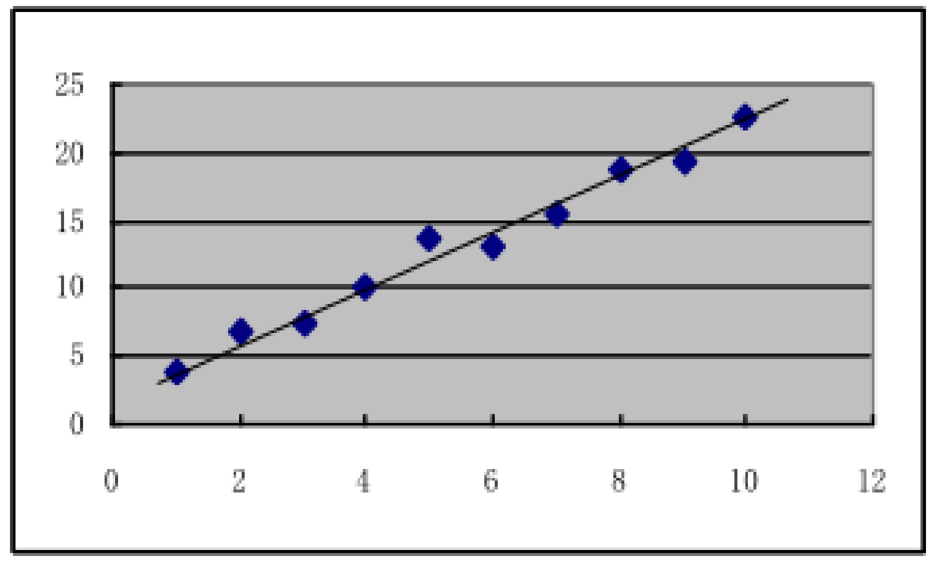
- 函数关系

## 非确定关系

- 线性关系（相关关系）
- 非线性关系




函数关系




线性（正）相关

# 线性回归

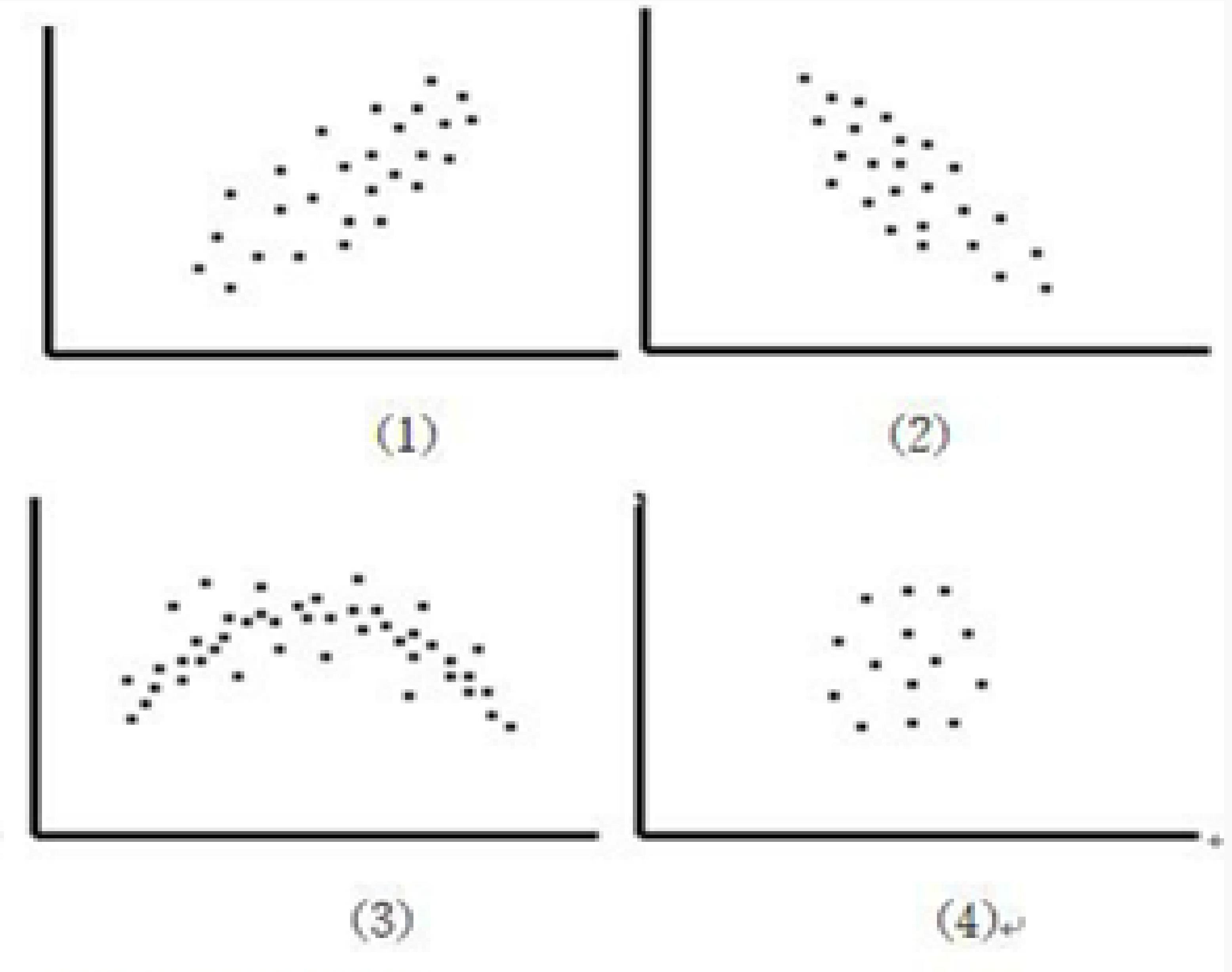
散点图观察相关性

相关性的程度

- 相关系数

**相关关系不是因果关系**

**独立和不相关**



## 协方差

- cov
- 协方差取值范围（-∞，+∞）

$$
\begin{aligned}
Cov(X,Y) &= E[(X-E[X])(Y-E[Y])] \\
&= E[XY]-2E[Y]E[X]+E[X]E[Y] \\
&= E[XY]-E[X]E[Y]
\end{aligned}
$$

## 相关系数

- cor
- 对协方差的单位化，取值范围[-1，+1]

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{D(X)}\sqrt{D(Y)}}$$

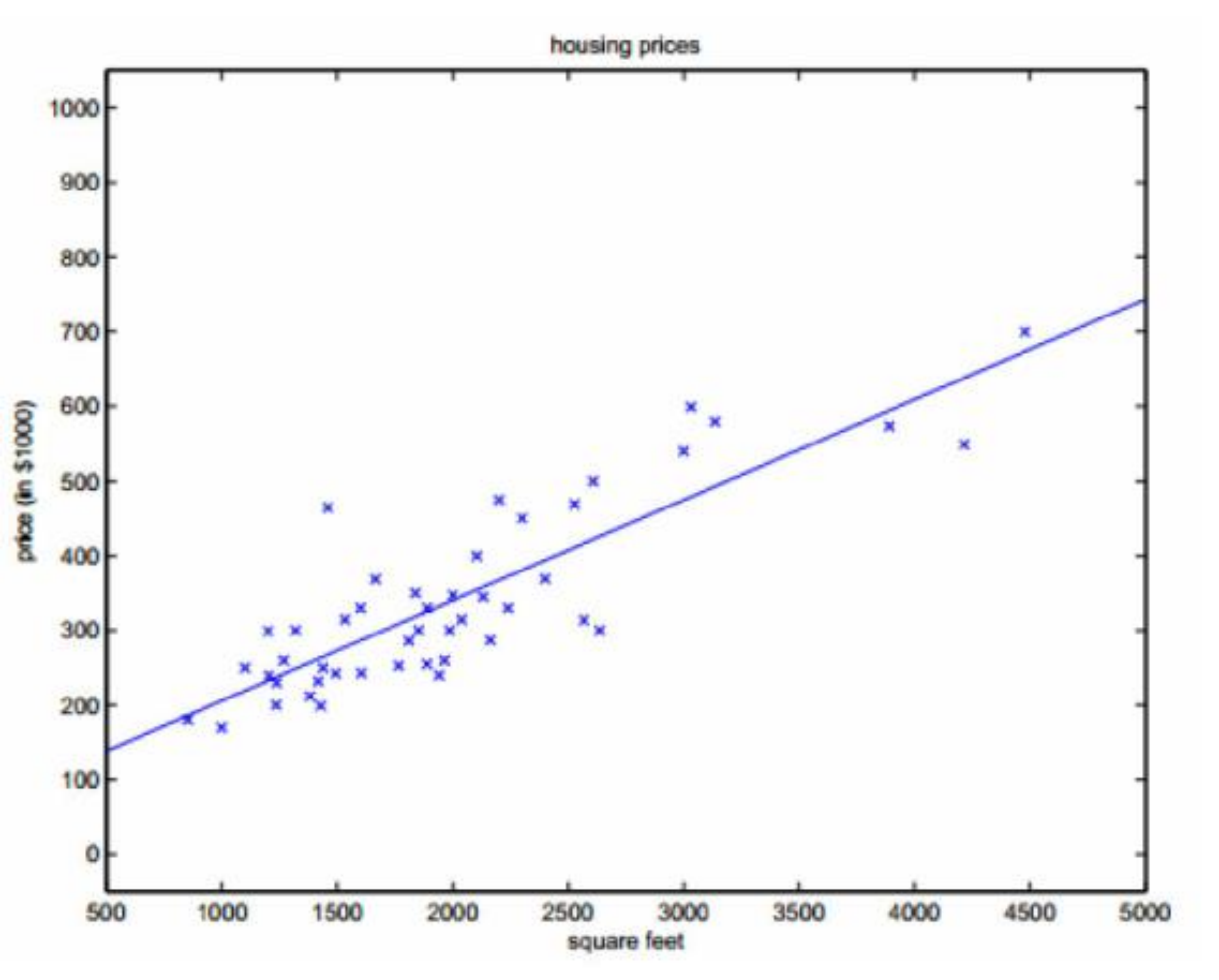
# 最小二乘法

# 线性回归

## 一元线性回归

$$Y = \beta_0 + \beta_1 X_1 + \varepsilon$$

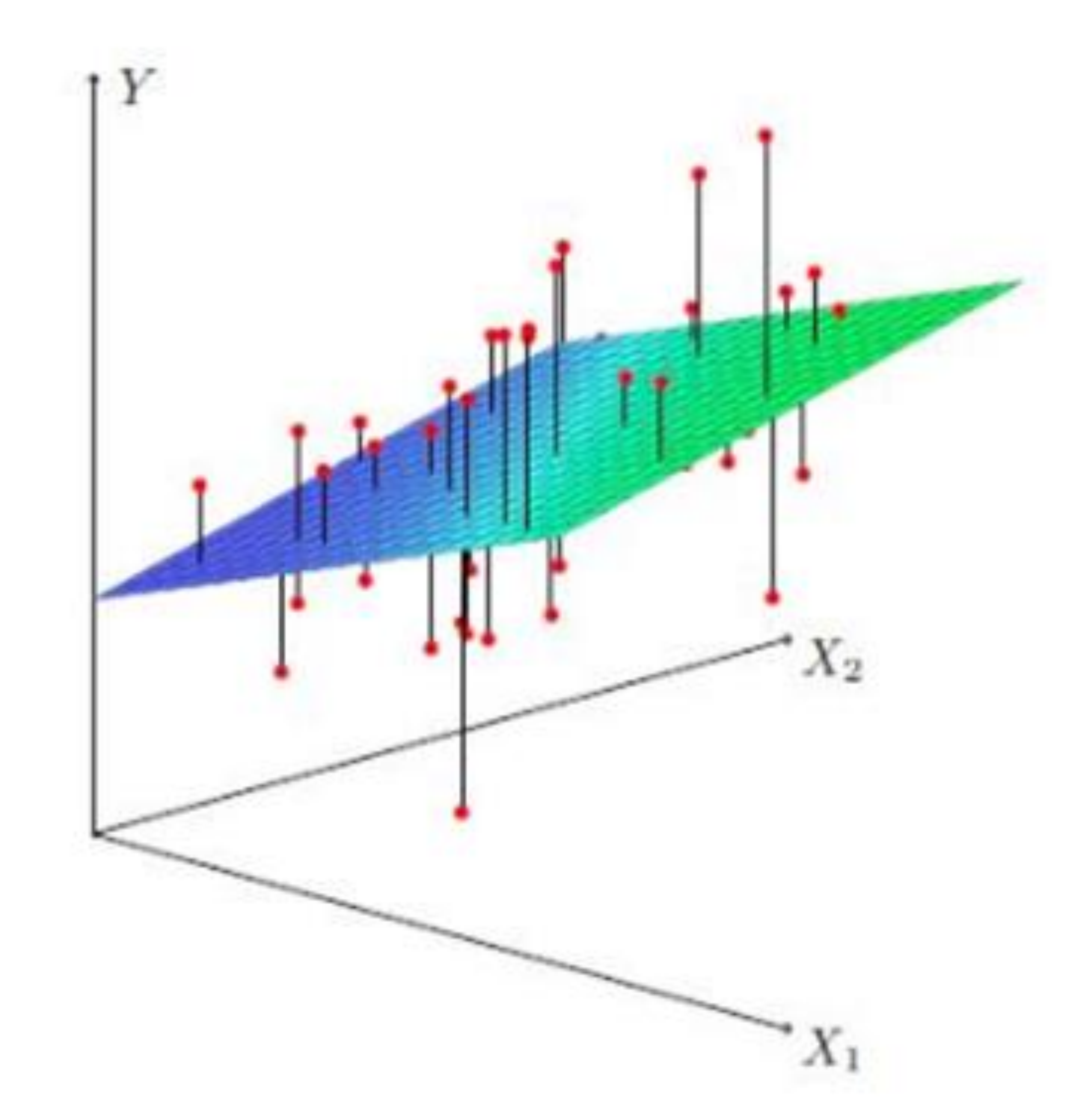
- ε代表残差（误差）



# 线性回归

## 多元线性回归

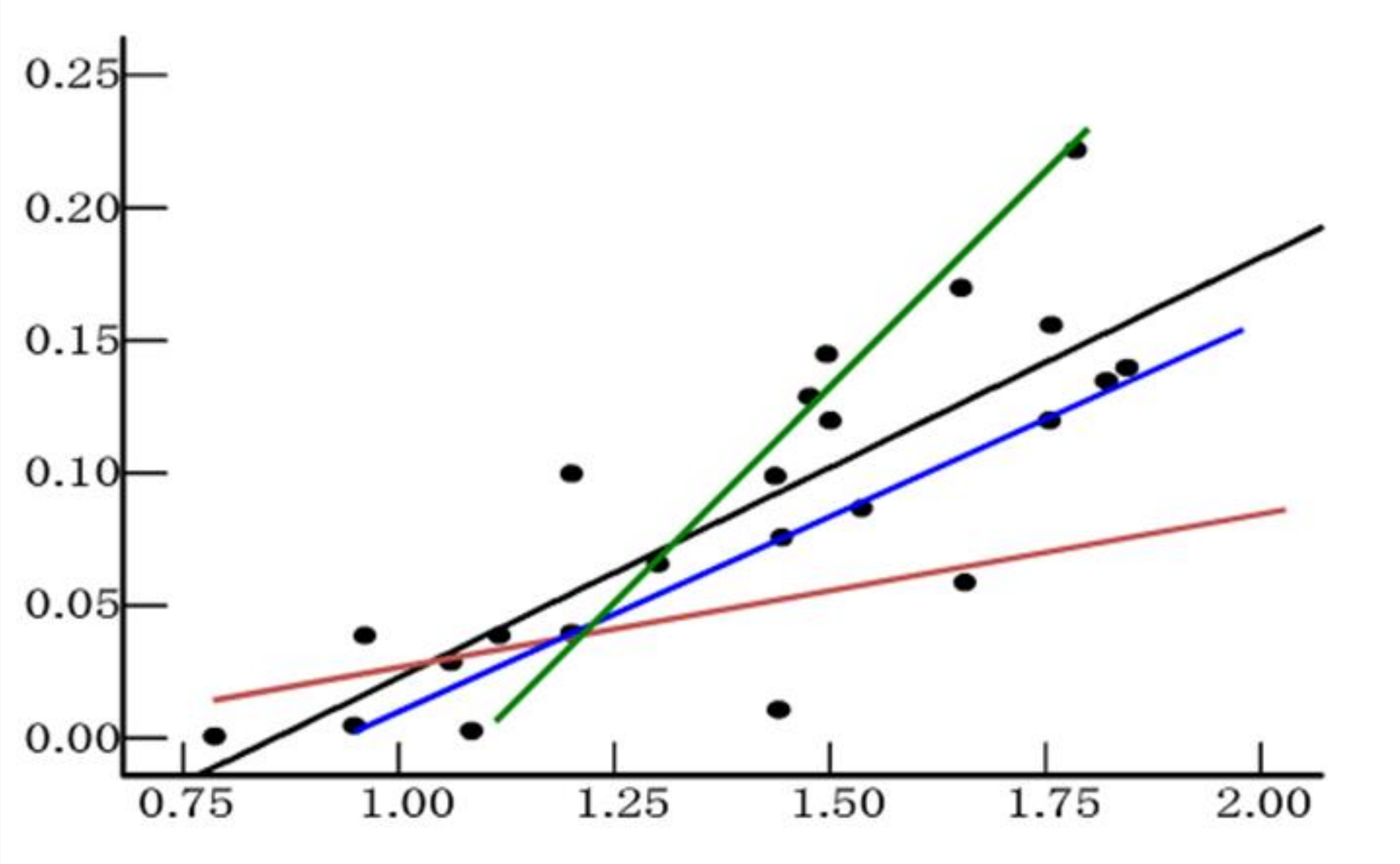
$$Y=\beta_0+\beta_1X_1+\beta_2X_2+\ldots\ldots+\beta_mX_m+\varepsilon$$



## 最佳回归线

不同的人会找到不同的“最佳”回归线

残差平方和最小为“最佳”



# 线性回归

## 最小二乘法（OLS）

- 残差平方和最小
- 残差就是y的实际值和y的回归值之间的差异，就是随机误差

## 目标函数

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(h_\theta(x^{(i)}) - y^{(i)})^2$$

# 梯度下降法

## 数值优化

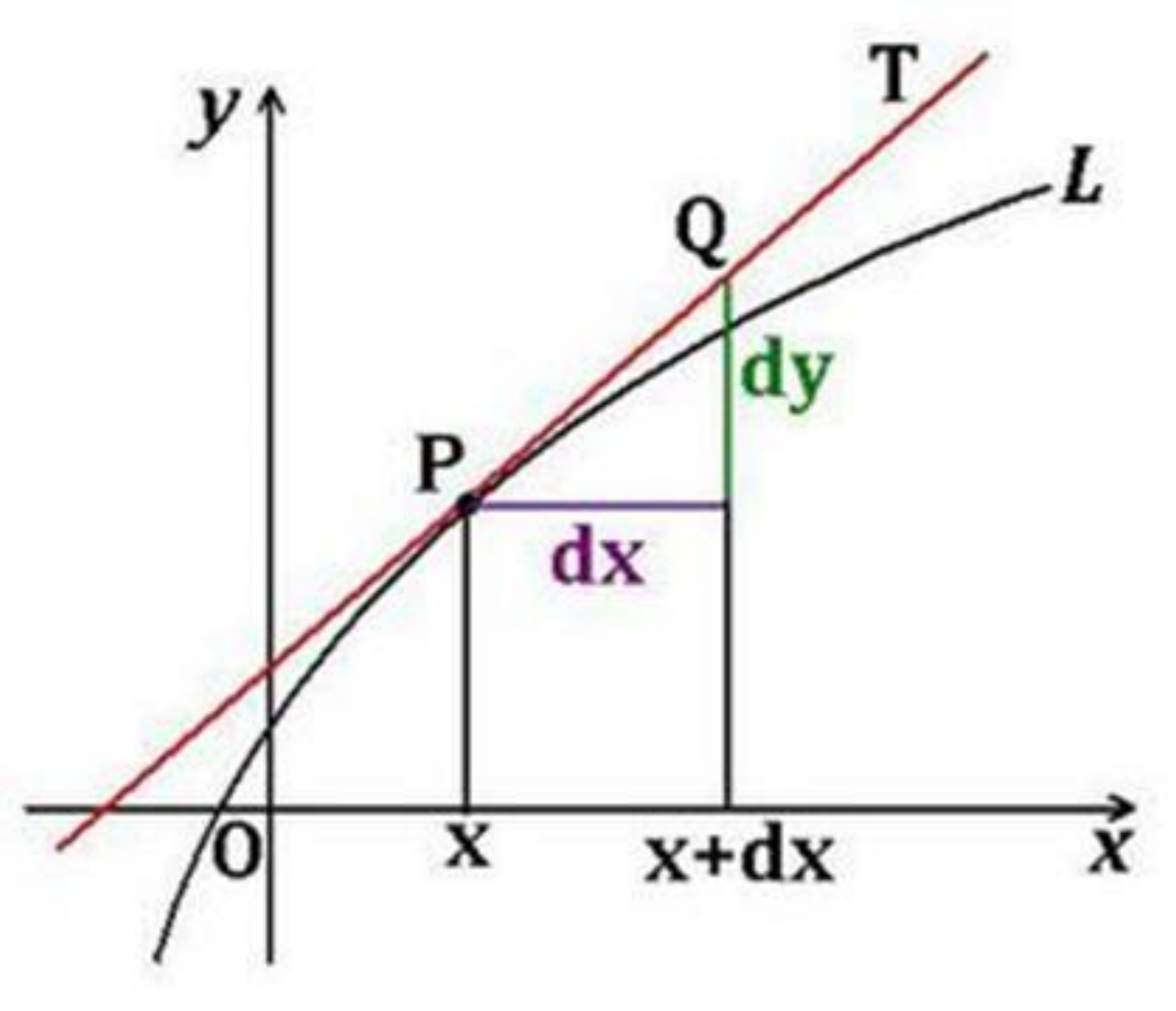
- 解析解
- 数值解
- 在机器学习中，基本上都是求数值解，只有线性回归能得到解析解

## 为什么要用梯度下降方法求解

- 绕过对于矩阵求逆
- 在计算机算法上会更容易实现

## 高等数学知识复习

- 微分
- 全微分和偏微分
- 梯度向量
- 方向导数
- **梯度方向是函数值变化最快的方向**
- 梯度方向指向函数值增大的方向，负梯度方向指向函数值减少的方向



## 梯度的推导

$$
\begin{aligned}
\frac{\partial}{\partial \theta_j} J(\theta) &= \frac{\partial}{\partial \theta_j} \frac{1}{2} \left(h_\theta(x) - y\right)^2 \\
&= 2 \cdot \frac{1}{2} \left(h_\theta(x) - y\right) \cdot \frac{\partial}{\partial \theta_j} \left(h_\theta(x) - y\right) \\
&= \left(h_\theta(x) - y\right) \cdot \frac{\partial}{\partial \theta_j} \left( \sum_{i=0}^{n} \theta_i x_i - y \right) \\
&= \left(h_\theta(x) - y\right) x_j
\end{aligned}
$$

## 梯度下降法伪代码

Repeat until convergence {

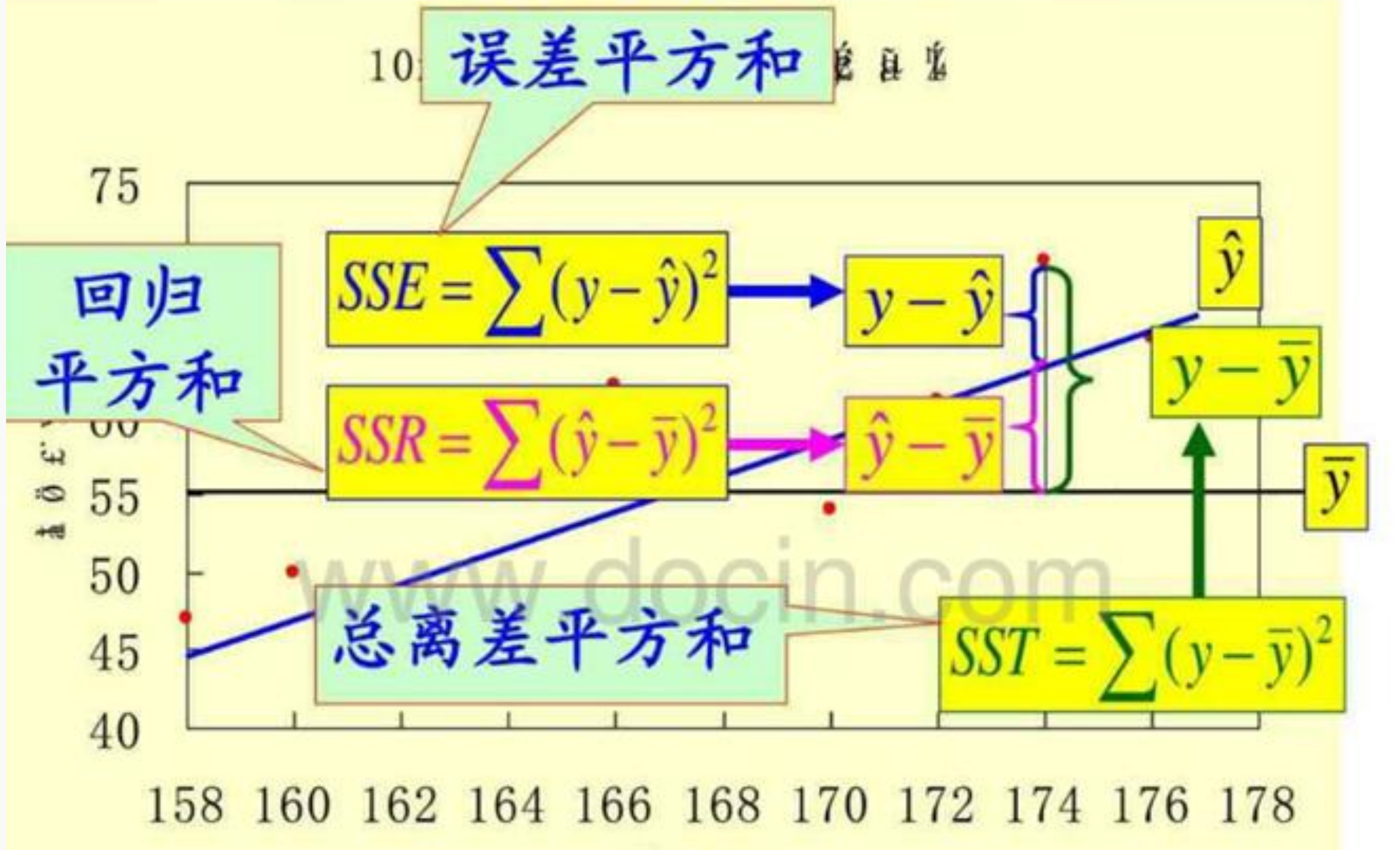
$$\theta_j := \theta_j + \alpha \sum_{i=1}^{m} \left(y^{(i)} - h_\theta(x^{(i)})\right) x_j^{(i)}$$

}

## α学习速率，超参数

## 决定系数 $R^2$

- 贡献有多大?
- 回归平方和占总平方和的比例
- 等于相关系数的平方



# 岭回归

# 线性回归

## 线性代数无法求解

- 变量之间存在线性关系
- 变量数量大于记录数

## 不存在广义逆

$$Y = X\beta + \varepsilon$$

$$\beta = (X^T X)^{-1} X^T y$$

## 引入扰动后求解

$$\hat{\beta}(k) = (X^{T}X + kI)^{-1}X^{T}y$$

等价于

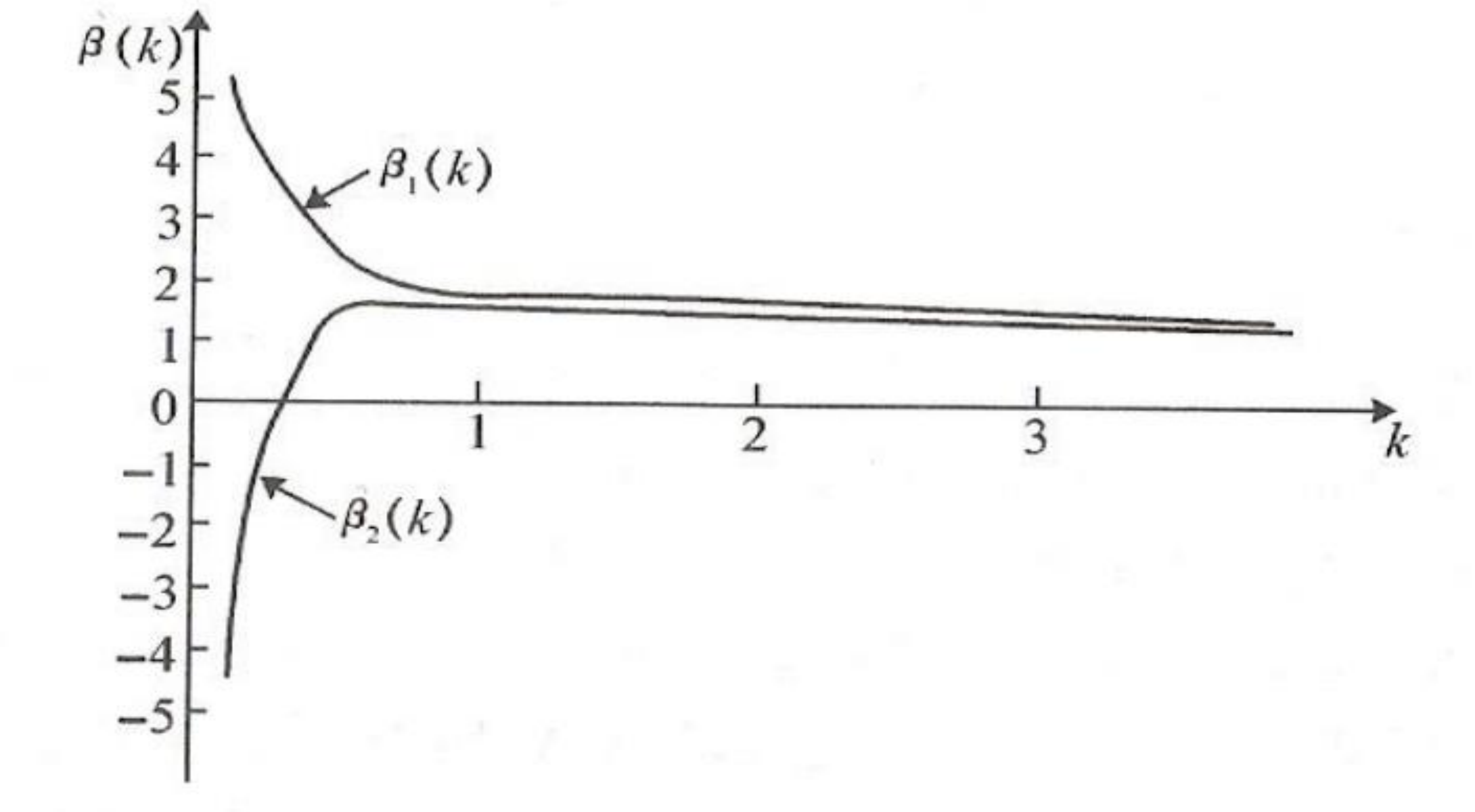
$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}}\left\{\sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2 + k\sum_{j=1}^{p}\beta_j^2\right\}$$

$$\hat{\beta}^{\text{ridge}} = \underset{\beta}{\operatorname{argmin}}\sum_{i=1}^{N}\left(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\right)^2,$$

$$\text{subject to } \sum_{j=1}^{p}\beta_j^2 \le t,$$

## 岭迹图

- 多重共线性
- 筛选k



# Lasso、弹性网

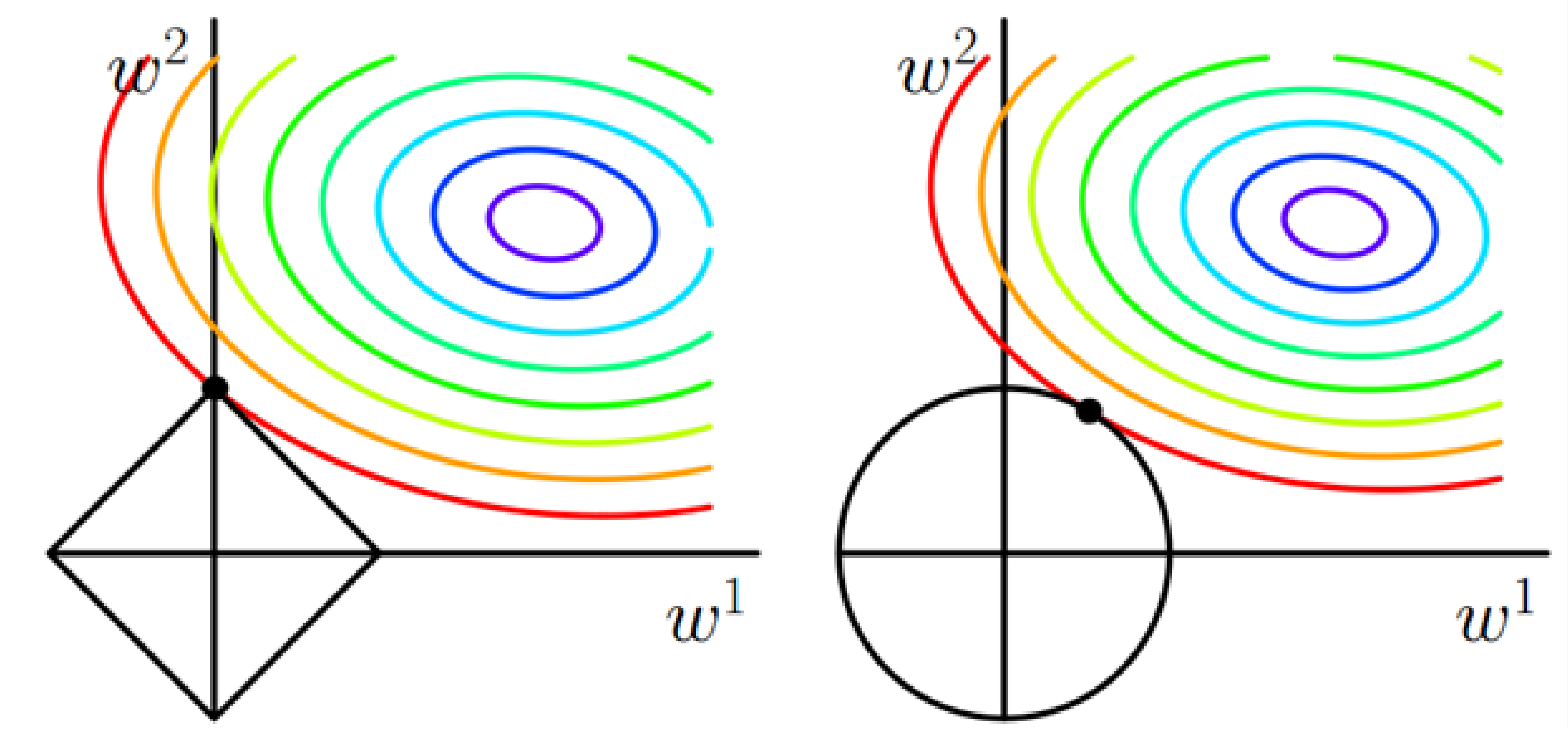
# 线性回归

## Lasso回归

- 岭回归的一般形式

$$\tilde{\beta} = \underset{\beta}{\operatorname{argmin}} \left\{ \sum_{i=1}^{N} \Big(y_i - \sum_{j=1}^{p} x_{ij}\beta_j\Big)^2 + k\sum_{j=1}^{p} |\beta_j| \right\}$$

- 变量的系数为0，实现筛选变量




(a) $\ell_1$-ball meets quadratic function. $\ell_1$-ball has corners. It's very likely that the meet-point is at one of the corners.

(b) $\ell_2$-ball meets quadratic function. $\ell_2$-ball has no corner. It is very unlikely that the meet-point is on any of axes.

## 弹性网

综合和岭回归和Lasso

$$\underset{\beta}{\operatorname{argmin}}\left\{\sum_{i=1}^{N}\left(y_i-\sum_{j=1}^{p}x_{ij}\beta_j\right)^2+\mathbf{k}\sum_{j=1}^{p}\left(\alpha\beta_j^2+(1-\alpha)|\beta_j|\right)\right\}$$