

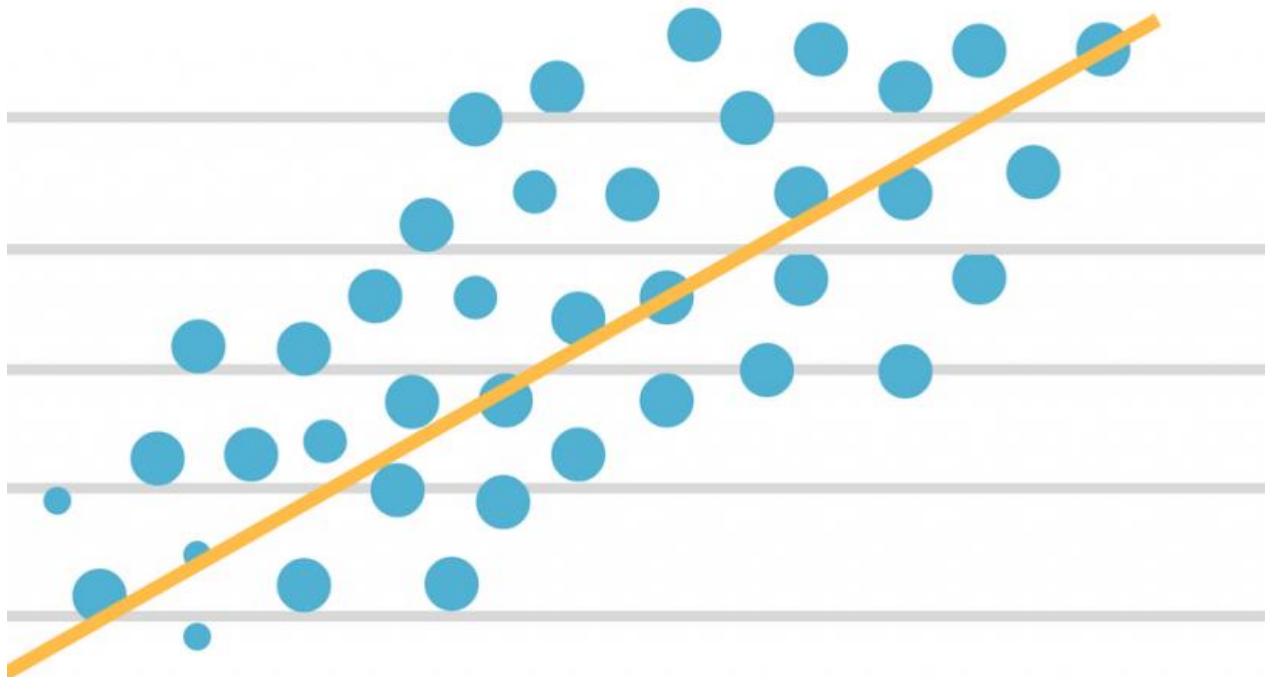
# 法律声明

---

- 本课件包括：演示文稿，示例，代码，题库，视频和声音等，小象学院拥有完全知识产权的权利；只限于善意学习者在本课程使用，不得在课程范围外向任何第三方散播。任何其他人或机构不得盗版、复制、仿造其中的创意，我们将保留一切通过法律手段追究违反者的权利。



关注 小象学院



# 回归模型

--Robin

# 目录

---

- 简单线性回归
- 多元线性回归

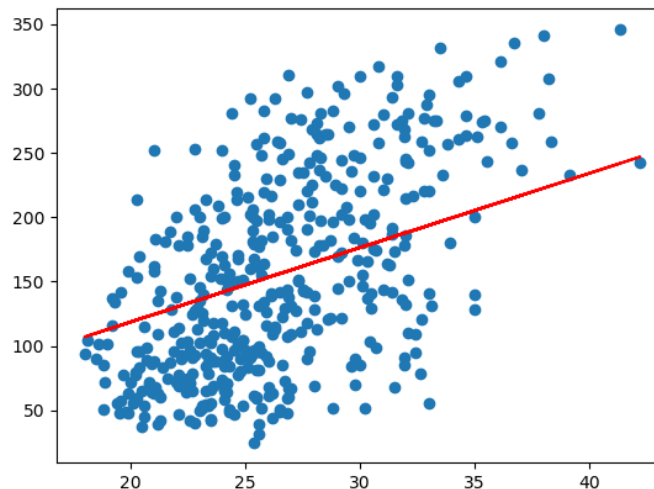
# 多元线性回归

## 线型回归

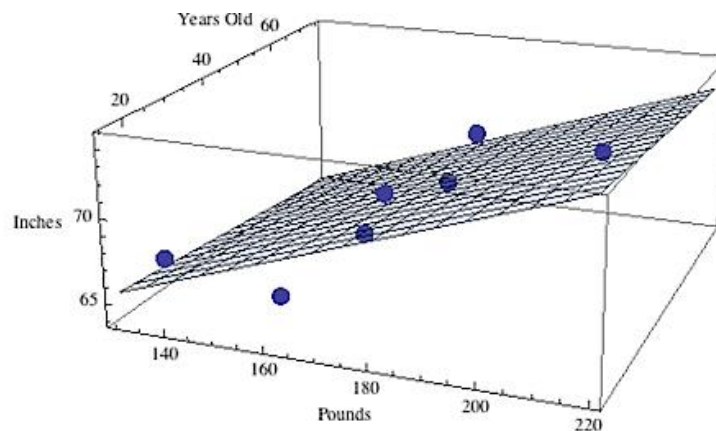
$$h_{\theta}(x) = \theta_0 + \theta_1 x_1 + \theta_2 x_2$$

$$h(x) = \sum_{i=0}^n \theta_i x_i = \theta^T x$$

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



低维



高维

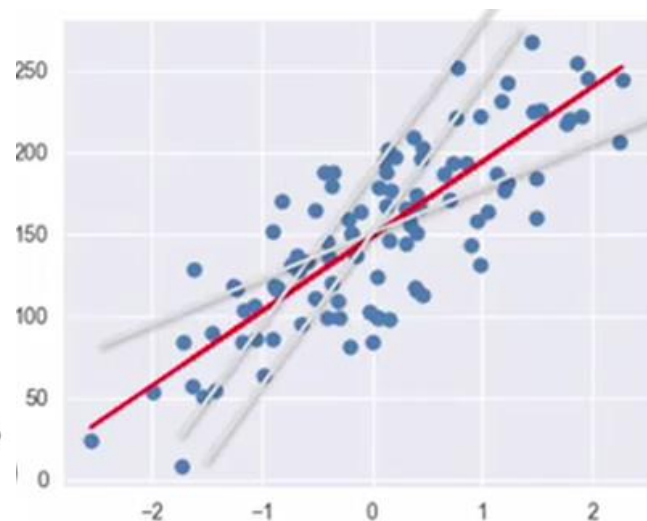
# 多元线性回归

## 线型回归

- 如何求解参数theta?

- 梯度下降法
- 最小二乘法 (Least Square)

$$\sum_{j=1}^n X_{ij}\beta_j = y_i, (i = 1, 2, \dots, m), \quad \mathbf{X}\boldsymbol{\beta} = \mathbf{y},$$



$$\mathbf{X} = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1n} \\ X_{21} & X_{22} & \cdots & X_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ X_{m1} & X_{m2} & \cdots & X_{mn} \end{bmatrix}, \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \quad \mathbf{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{bmatrix}$$

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} S(\boldsymbol{\beta}), \quad S(\boldsymbol{\beta}) = \sum_{i=1}^m \left| y_i - \sum_{j=1}^n X_{ij}\beta_j \right|^2 = \|\mathbf{y} - \mathbf{X}\boldsymbol{\beta}\|^2$$

对beta进行求导  
令导数为0

$$\Rightarrow (\mathbf{X}^T \mathbf{X}) \hat{\boldsymbol{\beta}} = \mathbf{X}^T \mathbf{y}. \quad \Rightarrow \hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y}$$

# 多元线性回归

---

## 线型回归

- 如何选择?
  1. 梯度下降法 (Gradient Descent)
  2. 最小二乘法 (Least Square)
- 数据量小的时候, 使用Least Square;
- 数据量大的时候, 可以考虑使用Gradient Descent

# 多元线性回归

## sklearn中调用线性回归

```
from sklearn.linear_model import LinearRegression

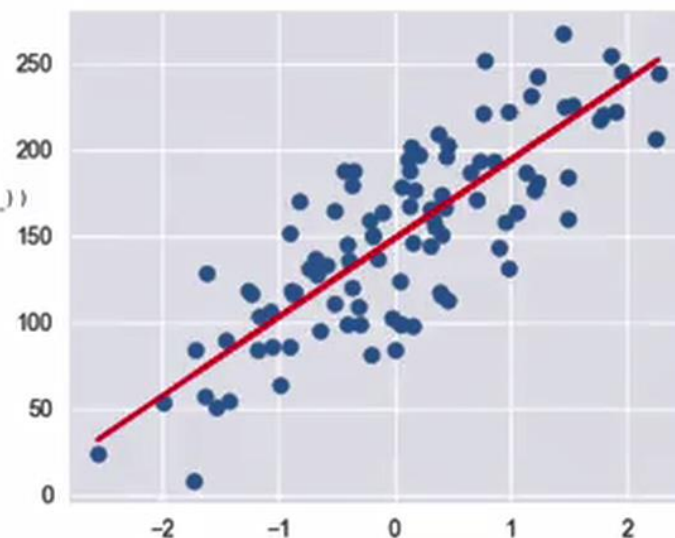
X_train, X_test, y_train, y_test =
    train_test_split(X_R1, y_R1, random_state = 0)

linreg = LinearRegression().fit(X_train, y_train)

print("linear model intercept (b): {}".format(linreg.intercept_))
print("linear model coeff (w): {}".format(linreg.coef_))
```

$$\hat{y} = \mathbf{w}_0 x_0 + b$$

Diagram illustrating the linear regression equation  $\hat{y} = \mathbf{w}_0 x_0 + b$ . The term  $\mathbf{w}_0$  is labeled as `linreg.coef_` and the term  $b$  is labeled as `linreg.intercept_`.



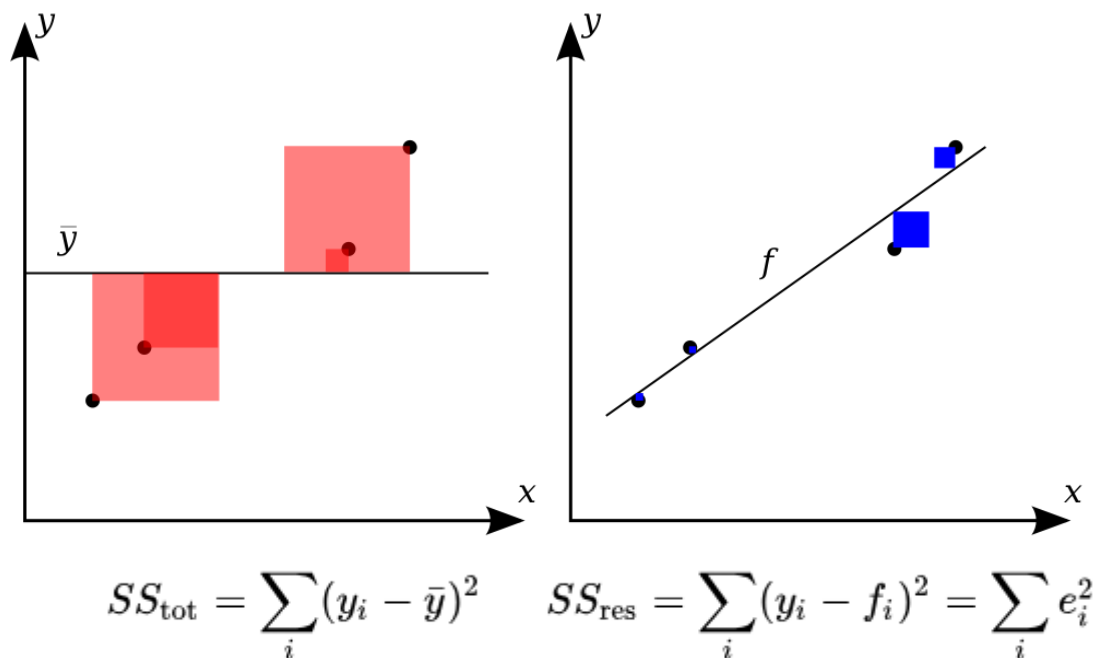
`sklearn.linear_model.LinearRegression`

# 多元线性回归

## 回归模型评价指标

- 决定系数(coefficient of determination,  $R^2$ , R squared)
- 反应了y的波动有多少百分比能被x的波动所描述
- 拟合程度越高，越接近于1

$$R^2 = 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}$$





# 联系我们

---

小象学院：互联网新技术在线教育领航者

— 微信公众号：**小象学院**

