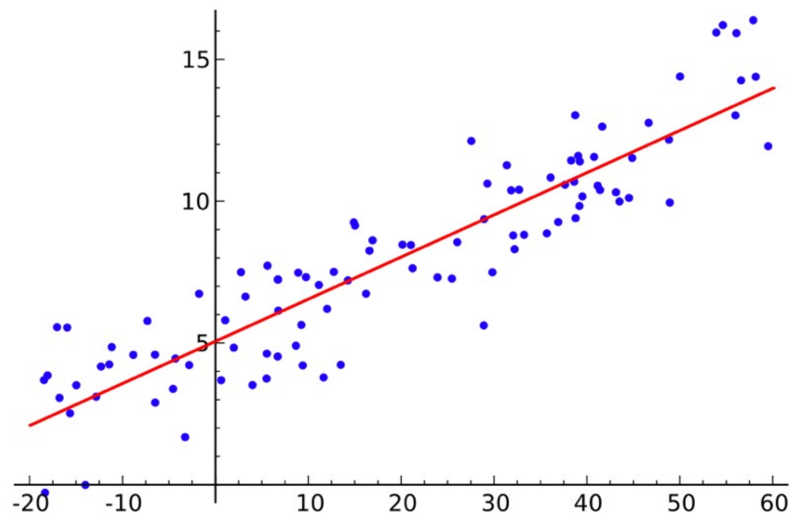
贪心科技
让每个人享受个性化教育服务

3. 线性回归

线性回归

- 线性回归（英语：linear regression）是利用线性回归方程的最小二乘函数对一个或多个自变量和因变量之间关系进行建模的方法



线性回归的应用

- 流行病学
  - 有关吸烟对死亡率和发病率影响的早期证据来自采用了回归分析的观察性研究
- 金融
  - 资本资产定价模型利用线性回归、分析和计算投资的系统风险。
- 经济学
  - 线性回归是经济学的主要实证工具。例如用来预测消费支出

线性回归的数学定义

- 数据 $(Y_i, X_{i1}, \ldots, X_{ip}), i = 1, \ldots, n$
- 模型 $Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \ldots + \beta_p X_{ip} + \varepsilon_i, \qquad i = 1, \ldots, n$

线性回归的数学定义(矩阵表达)

- 数据 $(Y_i, X_{i1}, \ldots, X_{ip}), i = 1, \ldots, n$

- 模型 $Y = X\beta + \varepsilon$

Y 是一个包括了观测值的列向量 $Y_1, \ldots, Y_n$

β 是一个包括了参数值的列向量 $\beta_0, \ldots, \beta_p$

$$X = \begin{pmatrix} 1 & x_{11} & \cdots & x_{1p} \\ 1 & x_{21} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \cdots & x_{np} \end{pmatrix}$$

线性回归的数学解析解（矩阵表达）

- 数据 $(Y_i, X_{i1}, \ldots, X_{ip}),\ i = 1, \ldots, n$
- 模型 $Y = X\beta + \varepsilon$
- 解析解 $\hat{\beta} = (X^T X)^{-1} X^T Y$

股价预测

- 预测A股上市公司5天后的收盘价

选择题

- 下面哪一个是线性回归的解析解:
- A. $\hat{\beta} = \left(X^T X\right)^T X^{-1} y$
- B. $\hat{\beta} = \left(X X^T X\right)^{-1} X^T y$
- C. $\hat{\beta} = \left(X X^T\right)^{-1} X^T y$
- D. $\hat{\beta} = \left(X^T X\right)^{-1} X^T y$

贪心科技　让每个人享受个性化教育服务

THANKS

线性回归解析解的推导

损失函数: $$J(\boldsymbol{\theta}) = \frac{1}{2n}\sum_{i=1}^{n}\left(h_{\boldsymbol{\theta}}\left(\boldsymbol{x}^{(i)}\right) - y^{(i)}\right)^2$$

$$= \frac{1}{2n}\sum_{i=1}^{n}\left(\boldsymbol{\theta}^\intercal \boldsymbol{x}^{(i)} - y^{(i)}\right)^2$$

$$= \frac{1}{2n}\underbrace{(\boldsymbol{X}\boldsymbol{\theta} - \boldsymbol{y})^\intercal}_{\mathbb{R}^{1\times n}}(\underbrace{\overset{\mathbb{R}^{n\times(d+1)}}{\boldsymbol{X}}\overset{\mathbb{R}^{(d+1)\times 1}}{\boldsymbol{\theta}}}_{\mathbb{R}^{n\times 1}} - \underbrace{\boldsymbol{y}}_{\mathbb{R}^{n\times 1}})$$

Let:

$$\boldsymbol{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

解析解

- 线性回归的损失函数为凸函数, 具有全局最优解. 对损失函数求导, 使得损失函数的导数为0时, θ的值就是解析解.

$$
\begin{aligned}
\mathcal{J}(\boldsymbol{\theta}) &= \frac{1}{2n}\left(\boldsymbol{X}\boldsymbol{\theta}-\boldsymbol{y}\right)^{\top}\left(\boldsymbol{X}\boldsymbol{\theta}-\boldsymbol{y}\right) \\
&\propto \boldsymbol{\theta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\theta} - \underbrace{\boxed{\boldsymbol{y}^{\top}\boldsymbol{X}\boldsymbol{\theta}}}_{1\times 1} - \underbrace{\boxed{\boldsymbol{\theta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{y}}}_{1\times 1} + \boldsymbol{y}^{\top}\boldsymbol{y} \\
&\propto \boldsymbol{\theta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\theta} - 2\boldsymbol{\theta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{y} + \boldsymbol{y}^{\top}\boldsymbol{y}
\end{aligned}
$$

$$
\frac{\partial}{\partial\boldsymbol{\theta}}J(\boldsymbol{\theta}) = 0
$$

$$
\begin{aligned}
\frac{\partial}{\partial\boldsymbol{\theta}}\left(\boldsymbol{\theta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{X}\boldsymbol{\theta} - 2\boldsymbol{\theta}^{\top}\boldsymbol{X}^{\top}\boldsymbol{y} + \cancel{\boldsymbol{y}^{\top}\boldsymbol{y}}\right) &= 0 \\
(\boldsymbol{X}^{\top}\boldsymbol{X})\boldsymbol{\theta} - \boldsymbol{X}^{\top}\boldsymbol{y} &= 0 \\
(\boldsymbol{X}^{\top}\boldsymbol{X})\boldsymbol{\theta} &= \boldsymbol{X}^{\top}\boldsymbol{y} \\
\boldsymbol{\theta} &= (\boldsymbol{X}^{\top}\boldsymbol{X})^{-1}\boldsymbol{X}^{\top}\boldsymbol{y}
\end{aligned}
$$

解析解

$$\boldsymbol{\theta} = (\boldsymbol{X}^\intercal \boldsymbol{X})^{-1} \boldsymbol{X}^\intercal \boldsymbol{y}$$

$$\boldsymbol{X} = \begin{bmatrix} 1 & x_1^{(1)} & \dots & x_d^{(1)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(i)} & \dots & x_d^{(i)} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_1^{(n)} & \dots & x_d^{(n)} \end{bmatrix} \qquad \boldsymbol{y} = \begin{bmatrix} y^{(1)} \\ y^{(2)} \\ \vdots \\ y^{(n)} \end{bmatrix}$$

如果$X^T X$不可逆, 需要使用如下对策:

1. 使用伪逆代替其逆矩阵, python: numpy.linalg.pinv( )
2. 去掉容余特征(线性相关的特征)
3. 去掉某些特征使得 d <= n