第八章回归分析

2021年10月19日　　15:28

目录：

正文：

1. 回归概述

   概念：

   回归分析（regression analysis)指的是确定两种或两种以上变量间相互依赖的定量关系的一种统计分析方法。

   分类：

   - 因变量的多少：简单回归分析、多重回归分析
   - 自变量的多少：一元回归、多元回归
   - 自变量和因变量之间的关系：线性回归分析、非线性回归分析

2. 应用案例

   详细见PDF课件

3. 线性回归

   - 一元线性回归：

     $$m(x) = \beta_0 + \beta_1 x$$

     RSS残差平方和最小

     $$R(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - (\beta_0 + \beta_1 x_i))^2$$

     - 让函数R的β0偏导为0可以得到：

       $$\beta_0 = \overline{y} - \beta_1 \overline{x}$$

     - 让β1偏导为0得到：

       $$\beta_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\overline{x}\overline{y}}{\sum_{i=1}^{n} x_i^2 - n\overline{x}^2}$$

   - 多元线性回归：

$$m(x_i) = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \cdots + \beta_p x_{id} = \beta_0 + \sum_{k=1}^{d} \beta_k x_{ik}$$

用矩阵表示：

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & \cdots & x_{1d} \\ 1 & x_{21} & \cdots & x_{2d} \\ \cdots & \cdots & \cdots & \cdots \\ 1 & x_{n2} & \cdots & x_{nd} \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \cdots \\ \beta_d \end{bmatrix}$$

则有：

$$R(\beta) = (Y - X\beta)^T (Y - X\beta)$$

让偏导为0，得到：

$$\frac{\partial R}{\partial \beta}\Big|_{=0} \Rightarrow \frac{\partial}{\partial \beta}\left(Y^T Y - Y^T X\beta - \beta^T X^T Y + \beta^T X^T X\beta\right)\Big|_{=0}$$

$$\Rightarrow X^T X\beta = X^T Y$$

$$\beta = \left(X^T X\right)^{-1} X^T Y$$

- 当 $(X^T X))^{-1}$不可逆时无法求出β
- $(X^T X)$ 越趋近于0，回归系数趋于无穷大，此时得到的回归系数无意义

- 多项式回归
  - 只用直线是不够的：
    - $m(x) = \beta_0 + \beta_1 x_1 + \beta_2 x_2^2 + \cdots + \beta_d x_d^d = \beta_0 + \sum_{k=1}^{d} \beta_k x_{ik}^k$
    - $R(\beta) = (Y - X\beta)^T (Y - X\beta)$

4. 非线性回归

5. logistic回归

6. 评估
   - 均方误差MSE：
   - 均方根误差RMSE
   - 平均绝对误差MAE
   - 总离差平方和SST
   - 回归残差平方和SSE：
   - 决定系数