

拟合优度与模型选择

司继春

2025年8月

拟合优度

- 拟合优度 (goodness of fit) 即使用 x 对 y 进行拟合的效果, 即被解释变量 y 有多少可以被解释变量 x 所解释。
- 在回归分析中, 最常用的拟合优度的度量为可决系数 (coefficient of determination), 或称 R^2 (R-quared)。

拟合优度的度量

- 在拟合或者预测的应用中，我们经常会关注 x 对 y 的解释能力问题。特别的，我们关注 y 的方差中有多少是可以被 x 解释的。
- 从总体的层面，回忆公式

$$\mathbb{V}(y) = \mathbb{V}[\mathbb{E}(y|x)] + \mathbb{E}[\mathbb{V}(y|x)]$$

其中：

- $\mathbb{E}(y|x)$ 为 y 的方差中可以被 x 所预测的部分
- 而 $\mathbb{V}(y|x) = \mathbb{V}(u|x)$ 为 y 中不可被 x 预测的部分的条件方差
- 因而 y 的总方差就被分解为可预测部分的方差与不可预测部分的方差，那么

$$\frac{\mathbb{V}[\mathbb{E}(y|x)]}{\mathbb{V}(y)} = 1 - \frac{\mathbb{E}[\mathbb{V}(y|x)]}{\mathbb{V}(y)}$$

就可以被解释为 y 的方差中有多少比例是可以被 x 所预测的。

总平方和的分解

- 从样本角度也可以类似分解
- 记 y 的方差的分子为总平方和 (total sum of squares) :

$$\text{TSS} = \sum_{i=1}^N (y_i - \bar{y})^2 = Y' M_0 Y$$

其中 $M_0 = I - \frac{1}{N} \iota \iota'$ 。

- 分解 (过程见讲义) :

$$\text{TSS} = \text{ESS} + \text{RSS}$$

其中:

$$\text{ESS} = \hat{Y}' M_0 \hat{Y} = \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$$

$$\text{RSS} = \hat{e}' \hat{e} = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

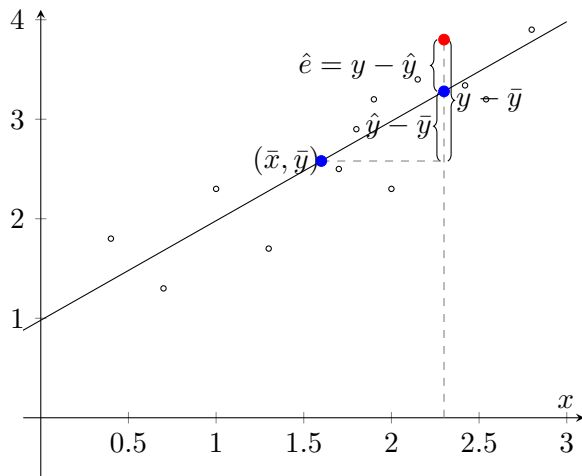
总平方和的分解

$$\text{ESS} = \hat{Y}' M_0 \hat{Y} = \sum_{i=1}^N (\hat{y}_i - \bar{\hat{y}})^2 = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2$$

$$\text{RSS} = \hat{e}' \hat{e} = \sum_{i=1}^N (y_i - \hat{y}_i)^2$$

- 回归平方和 (explained sum of squares) : x 可以解释的部分
- 残差平方和 (residual sum of squares) : x 不能解释的部分

总平方和的分解



R^2 的定义

定义：

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}$$

R^2 度量了使用 x 对 y 进行预测时， x 可以解释多少部分的 y 的方差。 R^2 一定是在 $[0, 1]$ 之间的。

- 当 $R^2 = 0$ 时，对应于 $\text{RSS} = \text{TSS}$ ，此时 $\hat{y}_i = \bar{y}$ ，也就是说不管有没有 x ，对 y 的最优预测都是 $\bar{y}$ ，意味着 x 对 y 完全没有任何解释能力。
- 而如果 $R^2 = 1$ ，此时 $\text{ESS} = \text{TSS}$ ，即 $\hat{y}_i = y_i$ ，达到了完美拟合。

R^2 的取值范围

- TSS完全可以看作是只包含常数项的RSS
- 如果回归中包含常数项，那么预测结果不可能比只有常数项更差，从而必然会有 $RSS < TSS$ 。
- 但是，如果回归中不包含常数项，得到的预测可能效果会非常差（比只用常数项预测的结果还差），从而可能会出现 $TSS < RSS$ 的情况，从而得到 $R^2 < 0$ 。
- 所以， $R^2 \in [0, 1]$ 的前提条件是回归中包含常数项！

变量增加时的 R^2

- 如果我们在回归中添加新的解释变量，由于出现了更多的信息，因而 R^2 值不会降低。
- 们假设 x_1 是一个 $N \times 1$ 维列向量，而 X_2 为 $N \times K$ 维其他解释变量，线性回归：

$$Y = \beta_1 x_1 + X_2 \beta_2 + e$$

的最小二乘估计分别为 $\hat{\beta}_1, \hat{\beta}_2$ 。

- 令残差： $\hat{e} = Y - \hat{\beta}_1 x_1 - X_2 \hat{\beta}_2$ ，那么拟合优度

$$R^2 = 1 - \frac{\hat{e}'\hat{e}}{\text{TSS}}$$

变量增加时的 R^2

- 如果只对 X_2 做回归，即：

$$Y = X_2\gamma + \epsilon$$

记其最小二乘估计为 $\hat{\gamma}$ ，残差为 $\hat{\epsilon} = Y - X_2\hat{\gamma}$ ，那么其 R^2 为：

$$R_2^2 = 1 - \frac{\hat{\epsilon}'\hat{\epsilon}}{\text{TSS}}$$

变量增加时的 R^2

定理

在上述回归模型中, $R_2^2 \leq R^2$ 。

目标函数为:

$$\min_{\beta_1, \beta_2} (Y - \beta_1 x_1 + X_2 \beta_2)' (Y - \beta_1 x_1 + X_2 \beta_2)$$

而如果只对 X_2 做回归, 目标函数为:

$$\min_{\gamma} (Y - X_2 \gamma)' (Y - X_2 \gamma)$$

相当于限制 $\beta_2 = \gamma, \beta_1 = 0$, 从而必然有: $\hat{e}'\hat{e} \leq \hat{\epsilon}'\hat{\epsilon}$

变量增加时的 R^2

定理

R^2 和 R_2^2 满足：

$$R^2 = R_2^2 + [\text{Corr}(\hat{\epsilon}, \hat{v})]^2 (1 - R_2^2)$$

其中 $\hat{\epsilon}$ 为回归 $x_1 = X_2\delta + v$ 的残差，即 $\hat{v} = x_1 - X_2\hat{\delta}$ 。

证明见讲义。其中 $\text{Corr}(\hat{\epsilon}, \hat{v})$ 度量了 Y 和 x_1 同时不能被 X_2 解释的部分相关系数，即排除 X_2 影响之后的相关系数，因而也被称为 Y 和 x_1 的偏相关系数（partial correlation）。