A Geometric Approach to Compare Variables in a Regression Model

Author(s): Johan Bring
Source: The American Statistician ,
Feb., 1996, Vol. 50, No. 1 (Feb., 1996), pp. 57-62

112354015 統碩一 林子涵

CONTENT

1. Introduction

- 為什麼幾何方法在迴歸分析中很少被使用？

1. 現今常用的代數方法已發展完全
2. Fisher、Durbin等著名的統計學家曾經使用，導致部分的人認為只有聰明的人才能駕馭幾何方法
3. 由於一般人對空間的認知只能到三維，當變數較多時難以用幾何方法詮釋

1. Introduction

- 研究目的：

1. 展現幾何方法的強大之處
2. 如何運用幾何方法選取變數與比較變數的重要性

2. A geometric presentation of multiple regression

現有相依變數y^*，p個解釋變數$x_1^*, \dots, x_p{}^*$，由於各變數的單位可能不同，所以先將以上變數都標準化，

可得$\|y\| = \|x_1\| = \cdots = \|x_p\| = 1$

再來配適迴歸模型，估計$\hat{y} = X\hat{\beta}$，

其中X：標準化解釋變數、$\hat{\beta}$：標準化迴歸係數

ŷ可視為y在X上的垂直投影

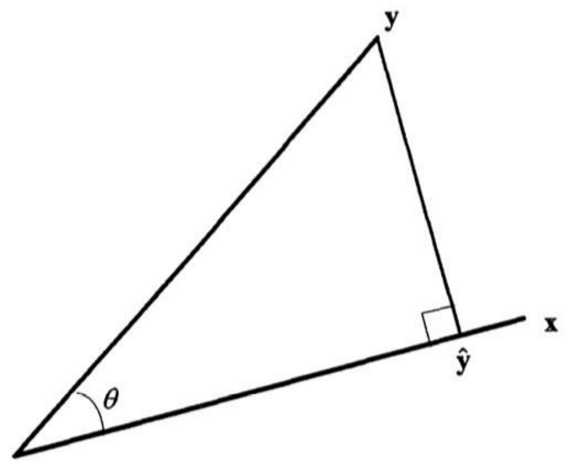


Figure 2. Linear Regression of y on x.

2. A geometric presentation of multiple regression

$$SST = \sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i)^2 = \|y\|^2 = 1$$

$$SSR = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2 = \sum_{i=1}^{n}(\hat{y}_i)^2 = \|\hat{y}\|^2$$

$$R^2 = \|\hat{y}\|^2$$

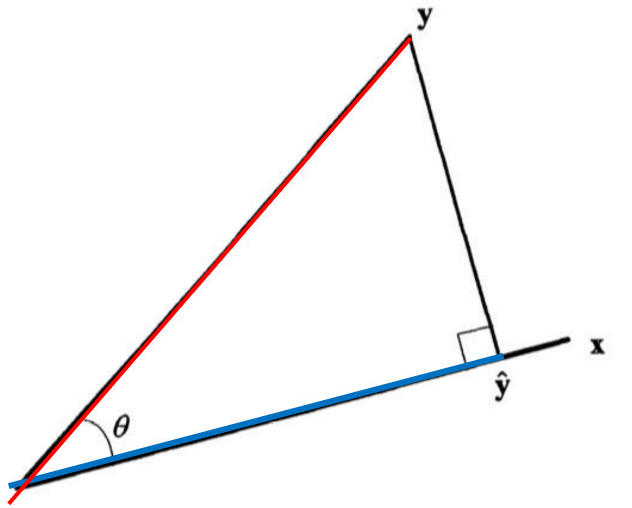


Figure 2. Linear Regression of y on x.

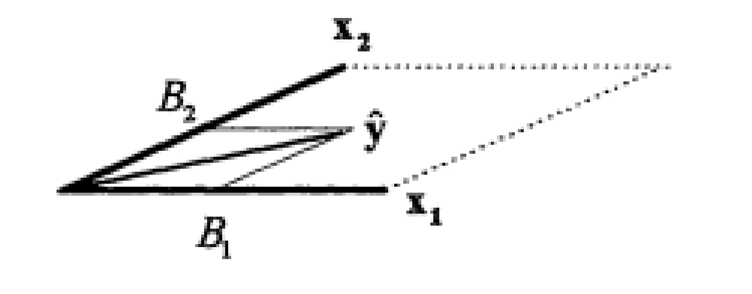
3. Standardized regression coefficients β_i

標準化迴歸係數估計方法：$\hat{\beta}_i^* = r_{iy}$

假設x_1，…，x_p彼此不相關，則其展開的空間為直角坐標系，

可得$\hat{y} = \hat{\beta}_1^* x_1 + \cdots + \hat{\beta}_p^* x_p$，因此$\hat{y}$可視為X空間上的向量，

且可用$\|\hat{y}\|$來決定「x_i對y的相對重要性」



4. Partitioning R^2

Method 1：Stepwise

1. 假設x_1，…，x_p彼此不相關，則每放入x_i可增加的R^2幅度為$\left(r_{iy}\right)^2$。
2. 假設x_1，…，x_p彼此相關，則R^2增加的幅度取決於變數放入順序。變數之間彼此的相關程度愈高，相對重要性的衡量愈取決於「變數納入順序」。

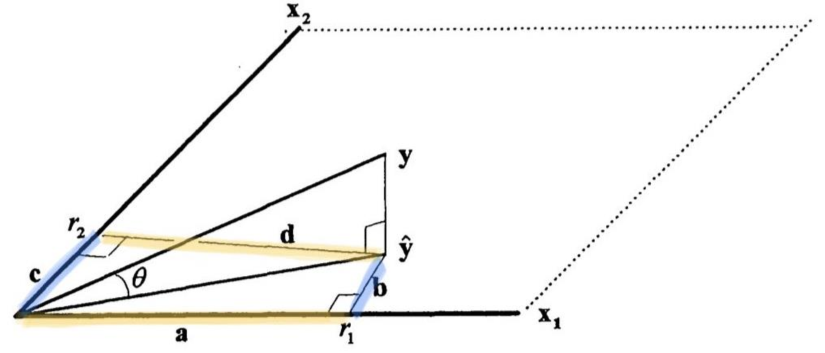


Figure 6. Selection of the First Variable by the Stepwise Regression Procedure.

Method 2：Average Stepwise

直接考慮所有可能變數的順序組合，將「納入x_i時R^2增加的平均值」視為x_i貢獻程度。

當p=2時：

$$x_1\text{的貢獻} = \frac{1}{2}(\|a\|^2 + \|d\|^2)$$

$$x_2\text{的貢獻} = \frac{1}{2}(\|b\|^2 + \|c\|^2)$$

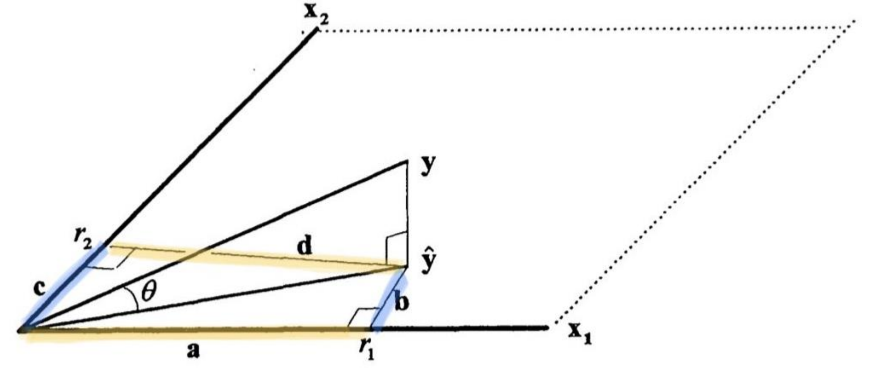


Figure 6. Selection of the First Variable by the Stepwise Regression Procedure.

Method 3：The Product Measure

$$R^2 = \hat{\beta}_1^* r_{1y} + \hat{\beta}_2^* r_{2y} + \cdots + \hat{\beta}_p^* r_{py}$$

其中，$\hat{\beta}_i^*$是標準化迴歸係數，r_{iy}是x_i和y的相關係數，

可使用$\hat{\beta}_i^*\, r_{iy}$來衡量x_i對R^2的貢獻。

5. Conclusion

- 以幾何的角度來詮釋迴歸分析
- 以標準化迴歸係數、判定係數等指標來選取變數
- 由於幾何方法較為抽象，因此在舉例時應用較簡單的方式說明

THANKS