Спецкурс "Регрессия"

Шкляев А.В.

15 января 2019 г.

0.1 Гипотеза в параметрическом виде

Предположим, что в линейной модели $\vec{Y} = X\vec{a} + \vec{\varepsilon}$ мы проверяем гипотезу $H_0 : C\vec{a} = \vec{r}$, где C — заданная матрица $m \times k$ полного ранга, $k \leq m$, $\vec{r}$ — заданный вектор.

Распишем разность

$$||\vec{Y} - X\vec{a}||^2 = ||\vec{Y} - X\widetilde{a}||^2 + 2\langle \vec{Y} - X\widetilde{a}, X(\widetilde{a} - \vec{a})\rangle + ||X(\widetilde{a} - \vec{a})||^2,$$

где $\vec{a}$ — произвольный вектор, такой что $C\vec{a} = \vec{r}$, а $C\widetilde{a} = \vec{r}$ — искомый минимум. Тогда нам достаточно подобрать такое $\widetilde{a}$, что $\langle X^t(\vec{Y} - X\widetilde{a}), \vec{b}\rangle = 0$ при всех $\vec{b}$: $C\vec{b} = 0$. При этом $X^t\vec{Y} - X^tX\widehat{a} = 0$, где $\widehat{a}$ — оценка МНК для общей задачи, откуда вектор $X^tX(\widehat{a} - \widetilde{a})$ должен быть ортогонален всем векторам $\vec{b} : C\vec{b} = 0$. Естественно потребовать

$$X^tX(\widehat{a} - \widetilde{a}) = C^t\vec{u}$$

при некотором $\vec{u}$. Тогда

$$(\widehat{a} - \widetilde{a}) = (X^tX)^{-1}C^t\vec{u},\ C(\widehat{a} - \widetilde{a}) = C(X^tX)^{-1}C^t\vec{u} = C\widehat{a} - \vec{r}.$$

Отсюда

$$\vec{u} = C(X^tX)^{-1}C^t)^{-1}(C\widehat{a} - \vec{r}),\ \widetilde{a} = \widehat{a} - (X^tX)^{-1}C^t(C(X^tX)^{-1}C^t)^{-1}(C\widehat{a} - \vec{r}) \tag{1}$$

При этом

$$||\vec{Y} - X\widetilde{a}||^2 = ||\vec{Y} - X\widehat{a}||^2 + ||X(\widehat{a} - \widetilde{a})||^2, \tag{2}$$

где

$$||X(\widehat{a} - \widetilde{a})||^2 = (\widehat{a} - \widetilde{a})^tX^tX(\widehat{a} - \widetilde{a}) = (C\widehat{a} - \vec{r})^t(C(X^tX)^{-1}C^t)^{-1}(C\widehat{a} - \vec{r}) \tag{3}$$

где в последнем равенстве мы воспользовались формулой (1).

Таким образом, оценка в рассматриваемой задаче принимает вид (1) (ее также иногда называют обобщенной оценкой наименьших квадратов), а остаточная сумма квадратов для этой оценки имеет вид (2).

0.2 Снижение размерности

Предположим, что мы рассматриваем гипотезу $H_0 : \vec{a}_1 = ... = \vec{a}_k = 0$ о том, что часть коэффициентов можно убрать. Тогда

$$C = \begin{pmatrix} 1 & 0 & ... & 0 & 0 & ... & 0 \\ 0 & 1 & ... & 0 & 0 & ... & 0 \\ & & ... & & & & \\ 0 & 0 & ... & 1 & 0 & ... & 0 \end{pmatrix},$$

$\vec{r} = (0, ..., 0)$. Следовательно, для расстояния до пространства, заданного при гипотезе, можно записать выражение.

$$||X(\widehat{a} - \widetilde{a})||^2 = \widehat{a}^t C^t (C^t (X^t X)^{-1} C^t)^{-1} C \widehat{a}$$

Матрица $C^t (X^t X)^{-1} C$ представляет собой квадратную матрицу $k \times k$, представляющей верхний левый угол матрицы $(X^t X)^{-1}$. Для ее обращения мы воспользуемся следующей леммой

Лемма 1. *Пусть*

$$D = \begin{pmatrix} G & B \\ B^t & F \end{pmatrix},$$

где G, F — симметричные матрицы $k \times k$ и $(m-k) \times (m-k)$, B — матрица $k \times (m-k)$. Тогда

$$D^{-1} = \begin{pmatrix} H^{-1} & C, \\ C^t & K \end{pmatrix},$$

где $H = G - BF^{-1}B^t$, K — некоторая симметрическая матрица $(m-k) \times (m-k)$, C — некоторая матрица $k \times (m-k)$.

Доказательство. По существу вопрос только в том, почему верхний угол L матрицы D^{-1} имеет такой вид. Запишем условие $DD^{-1} = E$:

$$GL + BC^t = E_k,\ GC + BK = 0,\ B^t L + FC^t = 0,\ B^t C + FK = E_{m-k}.$$

Отсюда $C^t = -F^{-1}B^t L$, $GL - BF^{-1}B^t L = HL = E_k$, откуда $L = H^{-1}$, что и требовалось доказать. □

Замечание 1. *Можно указать явный вид $C = -H^{-1}BF^{-1}$, $K = F^{-1}(E - B^t H^{-1} B F^{-1})$ и получить предыдущее утверждение простым перемножением матриц.*

Таким образом, $C^t (X^t X)^{-1} C = H^{-1}$, откуда

$$H = (C^t (X^t X)^{-1} C)^{-1} = (G - BF^{-1}B^t)^{-1},$$

где $G = X_1^t X_1$, $B = X_1^t X_2$, $F = X_2^t X_2$, где $X_1 = (X_{\cdot,1}, ..., X_{\cdot,k})$, $X_2 = (X_{\cdot,k+1}, ..., X_{\cdot,m})$. Следовательно,

$$||X(\widehat{a} - \widetilde{a})||^2 = \widehat{a}_{1:k}^t (G - BF^{-1}B^t) \widehat{a}_{1:k} = \widehat{a}_{1:k}^t X_1^t (E - X_2 (X_2^t X_2)^{-1} X_2^t) X_1 \widehat{a}_{1:k},$$

где $\widehat{a}_{1:k} = (\widehat{a}_1, ..., \widehat{a}_k)$. Таким образом,

$$D_{0,1} = \widehat{a}_{1:k}^t X_1^t X_1 \widehat{a}_{1:k} - \widehat{a}_{1:k}^t X_1^t X_2 (X_2^t X_2)^{-1} X_2^t X_1 \widehat{a}_{1:k},\ D_1 = \vec{Y}^t (E - X(X^t X)^{-1} X^t) \vec{Y}.$$

Таким образом, критерий приобретает вид

$$\frac{\left(\widehat{a}_{1:k}^t X_1^t X_1 \widehat{a}_{1:k} - \widehat{a}_{1:k}^t X_1^t X_2 (X_2^t X_2)^{-1} X_2^t X_1 \widehat{a}_{1:k}\right)/k}{\vec{Y}^t (E - X(X^t X)^{-1} X^t) \vec{Y}/(n-m)} > f_{1-\alpha},$$

где f — квантиль $F_{k,n-m}$.

В частности, удобный вид критерий приобретает в случае ортогональных блоков предикторов, когда X_1 ортогонально X_2. В этом случае $D_{0,1} = \widehat{a}_{1:k}^t X_1^t X_1 \widehat{a}_{1:k}$.

Еще одним удобным случаем является случай $k = 1$, когда

$$D_{0,1} = \widehat{a}_1^2 \left(\sum_{i=1}^n X_{1,i}^2 - X_1^t X_2 (X_2^t X_2)^{-1} X_2^t X_1 \right).$$

При $k = m - 1$

$$D_{0,1} = \widehat{a}^t_{1:m-1} X^t_1 \left(E - \frac{1}{\sum_{l=1}^n x^2_{m,l}} (x_{m,i} x_{m,j}, i, j \leq n) \right) X_1 \widehat{a}_{1:m-1},\ R^2 = \frac{D_{0,1}}{D_0} = \frac{D_{0,1}}{\sum_{i=1}^n (y_i - \overline{y})^2}.$$

0.3 Последовательный спуск

Можно представлять себе регрессию следующим образом

1. Берем $Z_{\cdot,1} = X_{\cdot,1}$.
2. При каждом j осуществляем регрессию $X_{\cdot,j}$ на $Z_{\cdot,1}, ..., Z_{\cdot,j-1}$ и получаем коэффициенты

$$\widehat{\gamma}_{i,j} = \frac{\langle X_{\cdot,j}, Z_{\cdot,i} \rangle}{\langle Z_{\cdot,i}, Z_{\cdot,i} \rangle}$$

   и вектор остатков $Z_{\cdot,j} = X_{\cdot,j} - \sum_{i=1}^{j-1} \widehat{\gamma}_{i,j} Z_{\cdot,i}$.
3. Осуществляем регрессию y на z_p, находя коэффициент $\widehat{\beta}_m = \langle z_m, y \rangle / \langle z_m, z_m \rangle$.

При этом коэффициент $\widehat{\beta}_m$ совпадает с коэффициентом разложения Y по $X_{\cdot,m}$.