Неделя 1. МНК. Дополнительные материалы

Обозначения

Есть ряд математических понятий, которые в разных источниках имеют разные обозначения. Приведём здесь обозначения нашего курса максимально чётко, чтобы избежать путаницы в дальнейшем!

- $E(y_i)$ — математическое ожидание случайной величины y_i. В других источниках можно встретить обозначение $M(y_i)$.
- $Var(y_i)$ — дисперсия случайной величины y_i. В других источниках можно встретить обозначение $D(y_i)$, $V(y_i)$.
- $Cov(y_i, x_i)$ — ковариация случайных величин y_i и x_i, $Cov(y_i, x_i) = E(y_i x_i) - E(y_i)E(x_i)$.
- $Corr(y_i, x_i)$ — корреляция случайных величин y_i и x_i, $Corr(y_i, x_i) = Cov(y_i, x_i)/\sqrt{Var(y_i)Var(x_i)}$.
- $sCov(y, x)$ — выборочная (s — сокращение от sample) ковариация векторов $y_1, \ldots, y_n$ и $x_1, \ldots, x_n$, $sCov(y_i, x_i) = \sum(y_i - \bar{y})(x_i - \bar{x})/(n-1)$
- $sVar(y)$ — выборочная дисперсия вектора $y_1, \ldots, y_n$, $sVar(y_i, x_i) = \sum(y_i - \bar{y})^2/(n-1)$
- $sCorr(y, x)$ — выборочная корреляция векторов $y_1, \ldots, y_n$ и $x_1, \ldots, x_n$, $sCorr(y, x) = sCov(y_i, x_i)/\sqrt{sVar(y_i)sVar(x_i)}$.

Особую путаницу может вызвать обозначение RSS и ESS! Будьте очень бдительны! В данном курсе:

- RSS, сумма квадратов остатков, residual sum of squares, $RSS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$.
- ESS, объясненная сумма квадратов остатков, explained sum of squares, $ESS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$.

В других источниках можно столкнуться совершенно с противоположным обозначением (!!!), а именно, $RSS = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ (regression sum of squares, регрессионная сумма квадратов), $ESS = \sum_{i=1}^n (y_i - \hat{y}_i)^2$ (error sum of squares, сумма квадратов ошибок прогноза)

Меньше разногласий вызывают обозначения:

- Вектор зависимой переменной
$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}$$
- Вектор ошибок модели
$$\varepsilon = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}$$
- Матрица всех регрессоров (на примере двух объясняющих переменных):
$$X = \begin{pmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & z_n \end{pmatrix}$$

- Вектор прогнозных значений:

$$\hat{y} = \begin{pmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{pmatrix}$$

- Знакомьтесь, вектор из единичек

$$\vec{1} = \begin{pmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

- Вектор остатков, $\hat{\varepsilon} = y - \hat{y}$,

$$\hat{\varepsilon} = \begin{pmatrix} \hat{\varepsilon}_1 \\ \hat{\varepsilon}_2 \\ \vdots \\ \hat{\varepsilon}_n \end{pmatrix}$$

- Вектор неизвестных коэффициентов

$$\beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{pmatrix}$$

- Оценки неизвестных коэффициентов

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_k \end{pmatrix}$$

- Транспонирование матрицы мы будем обозначать штрихом, A'. В других источниках можно встретить обозначение A^T.

Модель в скалярной записи:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$$

Оцененная регрессия:

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 z_i$$

Модель в матричном виде:

$$y = X\beta + \varepsilon$$

Оценённая регрессия в матричном виде:

$$\hat{y} = X\hat{\beta}$$

Оценки коэффициентов в матричном виде

На примере модели $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ мы установили, что остатки, $\hat{\varepsilon}_i$ ортогональны регрессорам:

$$\begin{cases} \vec{1} \perp \hat{\varepsilon} \\ x \perp \hat{\varepsilon} \\ z \perp \hat{\varepsilon} \end{cases}$$

Другими словами, столбцы матрицы регрессоров X ортогональны остаткам регрессии, вектору $\hat{\varepsilon}$:

$$X'\hat{\varepsilon} = 0$$

Заметим, что здесь 0 — это вектор размера $k \times 1$. Подставляем формулу для остатков, $\hat{\varepsilon} = y - X\hat{\beta}$:

$$X'(y - X\hat{\beta}) = 0$$

Раскрываем скобки и переносим в разные стороны уравнения:

$$X'y - X'X\hat{\beta} = 0$$

$$X'X\hat{\beta} = X'y$$

Матрица X' имеет размер $k \times n$, поэтому на неё сокращать нельзя. Хотя иногда хочется :) А вот обратная матрица к матрице $X'X$ существует, если среди столбцов X нет линейно зависимых и $n \geq k$. Домножаем обе части уравнения слева на $(X'X)^{-1}$:

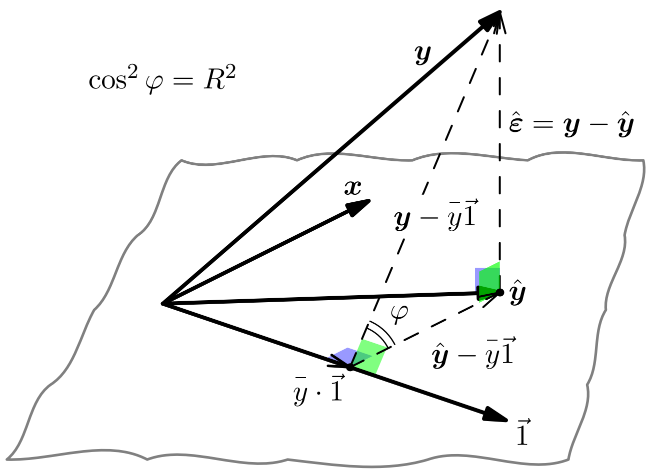
$$\hat{\beta} = (X'X)^{-1}X'y$$

Ура! Мы получили формулу для МНК-оценок множественной регрессии! Заметьте, что она подозрительно похожа на формулу МНК-оценки для случая одного оцениваемого параметра. В модели $y_i = \beta x_i + \varepsilon_i$ МНК-оценка коэффициента β будет иметь вид

$$\hat{\beta} = \frac{\sum x_i y_i}{\sum x_i^2} = \frac{1}{\sum x_i^2} \sum x_i y_i$$

Матрица-шляпница

Если $\hat{\beta} = (X'X)^{-1}X'y$, то вектор прогнозов, $\hat{y}$, будет равен $\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y$. Матрицу $H = X(X'X)^{-1}X'$ по-английски называют “hat-matrix”, матрицей-шляпницей, потому, что она надевает на y шляпку: $\hat{y} = H \cdot y$. Умножение любого вектора на матрицу H проецирует этот вектор на пространство, порождаемое регрессорами.



Поскольку сами регрессоры уже лежат в этом пространстве, то $H \cdot X = X$. Матрица H идемпотентная, то есть возведенная в произвольную натуральную степень даст саму себя, $H^n = H$. В этом легко можно убедиться либо перемножив руками H на H,

$$H \cdot H = X(X'X)^{-1}X'X(X'X)^{-1}X' = X(X'X)^{-1}X' = H$$

либо из геометрических соображений: умножение на H несколько раз подряд, это проецирование результата проецирования. А проекция от проекции совпадает с проекцией.

Собственными числами матрицы H могут быть только нули или единицы. Действительно, при проецировании часть векторов сохраняются (те, что лежали в пространстве регрессоров), часть превращается в ноль (те, что были ортогональны пространству регрессоров), а все другие при проецировании меняют направление.

Ранг матрицы-шляпницы можно посчитать, воспользовавшись тем, что $rk(AB) = rk(BA)$:

$$rk(X(X'X)^{-1}X') = rk(X'X(X'X)^{-1}) = rk(I_{k\times k}) = k$$