Конспектировал: Всеволод Овечкин

1. МНК без теории вероятности

Регрессионный анализ используют в основном для:

1. описания зависимости между признаками
2. предсказания значений переменной

В этом семинаре речь пойдет о методе наименьших квадратов (МНК) без теории вероятности, решении задачи поиска коэффициентов различными способами и геометрической интерпретации МНК.

1.1. Пример выявления зависимости

Рассмотрим небольшой пример, в котором y - это количество составленных задач по эконометрике, x - это количество съеденных пирожков. Попробуем выяснить как зависят эти две величины.

y	x
10	5
14	6
5	2

Таблица 1: Зависимость количества составленных задач по эконометрике от количества съеденных пирожков.

Введем обозначения y_i - фактическое значение объясняемой переменной, $\hat{y}_i$ - предсказанное значение объясняемой переменной, $\hat{u}_i$ - ошибка предсказания.

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}; \hat{u}_i = y_i - \hat{y}_i; \hat{u} = \begin{bmatrix} \hat{u}_1 \\ \hat{u}_2 \\ \vdots \\ \hat{u}_n \end{bmatrix}$$

Предположим, что зависимость выглядит следующим образом

$$\hat{y}_i = \hat{\beta} x_i$$

Как найти $\hat{\beta}$?. МНК метод позволяет найти такую $\hat{\beta}$, чтобы сумма квадратов ошибок была минимальна. Математически эту идею можно выразить несколькими способами:

$$\sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \to \underset{\hat{\beta}}{min}$$

$$d^2(y, \hat{y}) \to \underset{\hat{\beta}}{min}$$

$$(y - \hat{y})^T (y - \hat{y}) \to \underset{\hat{\beta}}{min}$$

Решим задачу минимизации суммы:

$$Q(\hat{\beta}) = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}x_i\right)^2 = \sum_{i=1}^{n} \left(y_i - 2y_i\hat{\beta}x_i + (\hat{\beta}x_i)^2\right) \to \underset{\hat{\beta}}{min}$$

$$Q'(\hat{\beta}) = \sum_{i=1}^{n} \left(-2y_ix_i + 2\hat{\beta}x_i\right)$$

Найдем условие первого порядка:

$$\sum_{i=1}^{n} \left(-2y_ix_i + 2\hat{\beta}x_i\right) = 0$$

$$-2\sum_{i=1}^{n} y_ix_i + 2\hat{\beta}\sum_{i=1}^{n} x_i = 0$$

$$\hat{\beta} = \frac{\sum_{i=1}^{n} y_ix_i}{\sum_{i=1}^{n} x_i}$$

$$\hat{\beta} = \frac{50 + 84 + 10}{25 + 36 + 4} = 2.215$$

Таким образом уравнение, позволяющее описать линейную зависимость y от x, и минимизирующее сумму квадратов остатков, выглядит следующим образом:

$$\hat{y} = 2.215 * x$$

1.2. Матричное дифференцирование

Пусть есть:

$$F = \begin{bmatrix} x_1 + x_2 & 2x_1^2 \\ x_1x_2 & cos(x_2) \end{bmatrix}$$

Тогда:

$$dF = \begin{bmatrix} dx_1 + dx_2 & 4x_1dx_1 \\ x_1dx_1 + x_2dx_2 & -sin(x_2)dx_2 \end{bmatrix}$$

Скаляр	Матрица
$d(a+b) = da + db$	$d(A+B) = dA + dB$
$d(ab) = bda + adb$	$d(AB) = dAB + AdB$
$d(cb) = cdb, c = const$	$d(CB) = CdB, C = const$
	$d(A^T) = (dA)^T$

Таблица 2: Сравнение операций дифференцирования скаляра и матрицы

Задачу поиска $\hat{\beta}$ так же можно решить при помощи матричного дифференциала:

$$Q(\hat{\beta}) = (y - \hat{y})^T (y - \hat{y}) \to \underset{\hat{\beta}}{min}$$

$$
\begin{aligned}
dQ(\hat\beta) = d(y - x\hat\beta)^T(y - x\hat\beta) + (y - x\hat\beta)^T d(y - x\hat\beta) = \\
= (-xd\hat\beta)^T(y - x\hat\beta) + (y - x\hat\beta)^T(-xd\hat\beta) = \\
= -2(y - x\hat\beta)^T(xd\hat\beta)
\end{aligned}
$$

Найдем условие первого порядка. Так как для любого $d\hat\beta$ должно выполняться $dQ() = 0$, то

$$
\begin{gathered}
-2(y - x\hat\beta)^T x = 0 \\
x^T(y - x\hat\beta) = 0 \\
x^T y - x^T x\hat\beta = 0 \\
x^T x\hat\beta = x^T y \\
\hat\beta = \frac{x^T y}{x^T x}
\end{gathered}
$$

1.3. Упражнения на матричное дифференцирование

Найти dA^{-1}, зная

$$
I_{n,n} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}; dI = 0; dA
$$

Решение:

$$
\begin{gathered}
dI = d(AA^{-1}) = dAA^{-1} + AdA^{-1} = 0 \\
dAA^{-1} = -AdA^{-1} \\
dA{-}1 = -A^{-1}dA^{-1}
\end{gathered}
$$

2. МНК с несколькими регрессорами

Пусть имеются данные:

y	x	z
$\vdots$	$\vdots$	$\vdots$

Таблица 3: Случайные данные

Предположим, что мы хотим оценить уравнение регрессии:

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

Введем обозначения: k - количество коэффициентов, n - количество наблюдений. Тогда

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}; \hat{y} = \begin{bmatrix} \hat{y}_1 \\ \hat{y}_2 \\ \vdots \\ \hat{y}_n \end{bmatrix}; X = \begin{bmatrix} 1 & x_1 & z_1 \\ 1 & x_2 & z_2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & z_n \end{bmatrix}; \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{bmatrix}$$

Уравнение регрессии можно записать в виде:

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

Найдем методом МНК вектор $\hat{\beta}$:

$$\begin{aligned} dQ(\hat{\beta}) = d(y - X\hat{\beta})^T(y - X\hat{\beta}) + (y - X\hat{\beta})^T d(y - X\hat{\beta}) = \\ = (-Xd\hat{\beta})^T(y - X\hat{\beta}) + (y - X\hat{\beta})^T(-Xd\hat{\beta}) = \\ = -2(y - X\hat{\beta})^T(Xd\hat{\beta}) \end{aligned}$$

Найдем условие первого порядка. Так как для любого $d\hat{\beta}$ должно выполняться $dQ() = 0$, то

$$-2(y - X\hat{\beta})^T X = 0$$
$$X^T(y - X\hat{\beta}) = 0$$
$$X^T y - X^T X\hat{\beta} = 0$$
$$X^T X\hat{\beta} = X^T y$$

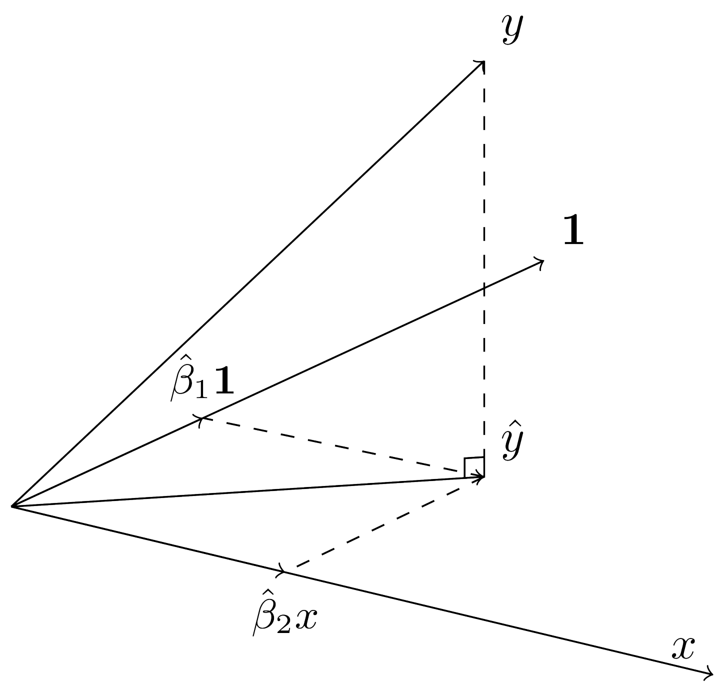
В случае если $X^T X$ - обратима и $det(X^T X) \neq 0$, то

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

3. Геометрическая интерпретация МНК

$$\hat{Y} = \hat{\beta}_1 \overline{1} + \hat{\beta}_2 x$$

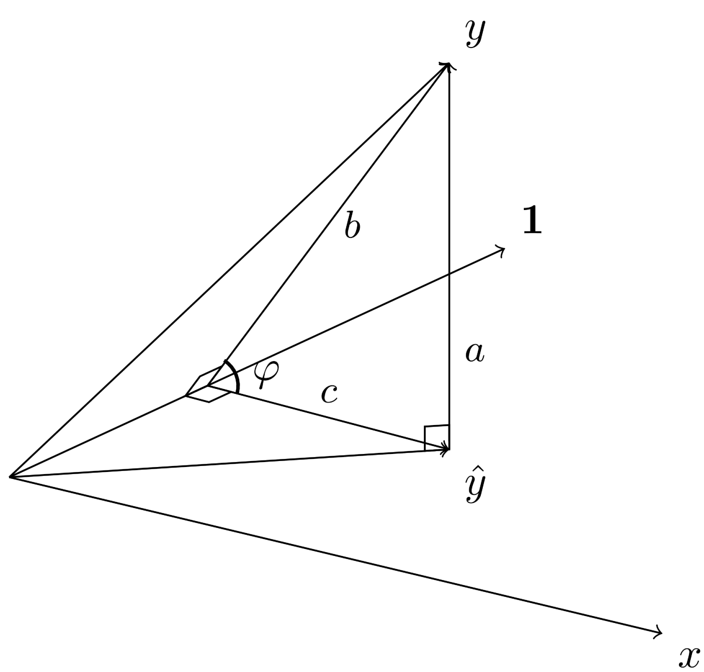
где, $\overline{1}$ - вектор столбец из единиц длинны n. Описанное выше можно переформулировать следующим образом $\hat{y} \in Lin(\overline{1}, x)$. Или $\hat{y}$ - проекция y на $Lin(\overline{1}, x)$



Так же из рисунка видно, что вертор ошибок $\hat{u}$ перпендикулярен $\hat{y}$ и подпространству $Lin(\overline{1}, x)$ Поэтому геометрический смысл условий первого порядка можно выразить следующим образом:

$$\begin{cases} \overline{1}\hat{u} = 0 \\ x^T\hat{u} = 0 \end{cases} \Leftrightarrow X^T\hat{u} = 0$$

4. Коэффициент детерминации. Геометрическая интерпретация



Возьмем тривиальную модель:

$$\hat{y} = \hat{\beta}_1 \overline{1}$$

В этой простой модели решение будет выглядеть следующим образом:

$$\hat{\beta} = \frac{\overline{1}^T y}{\overline{1}^T \overline{1}}$$

Тогда $\hat{y}$ можно будет найти по формуле:

$$\hat{y} = \overline{1} * \frac{\overline{1}^T y}{\overline{1}^T \overline{1}} = \overline{1} * \frac{\sum y_i}{n} = \overline{1} * \overline{y}$$

где, $\overline{y}$ - среднее значение y.

Вектор $\overline{1}\overline{y}$ на рисунке выше будет являть проекцией $\hat{y}$ на вектор $\overline{1}$. А так как $\hat{y}$ является проекцией y на пространство $Lin(\overline{1}, x)$, то вектор $\overline{1} * \overline{y}$ тоже будет являться проекцией y на вектор $\overline{1}$. Полученный треугольник с ребрами a, b, c является прямоугольным.

Предположим, что мы оцениваем какую-то линейную (не тривиальную) модель. Предсказанные значений y этой моделью обозначим как $\hat{y}$. Тогда стороны треугольника можно интерпретировать следующим образом:

- $||b||^2$ = $\sum_i (y_i - \overline{y})^2 = TSS$ (total sum of squares) - максимально возможная сумма квадратов ошибок модели при условии, что значения $\hat{y}$ прогнозируются тривиальной моделью;
- $||a||^2$ = $\sum_i (y_i - \hat{y}_i)^2 = RSS$ (residual sum of squares) - остаточная сумма квадратов, являющаяся суммой квадратов ошибок ($\hat{u}$) модели;
- $||c||^2$ = $\sum_i (\hat{y}_i - \overline{y})^2 = ESS$ (explained sum of squares) - объясненная сумма квадратов, показывающая часть TSS, которую удалось объяснить моделью;

По теореме Пифагора получаем: $TSS = ESS + RSS$. Тогда то, на сколько лучше наша модель предсказывает y по сравнению с тривиальной моделью можно измерить коэффициентом детерминации:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}; R^2 \in [0, 1]$$

Важное замечание: данная формула корректна в случае наличия $\overline{1}$ в модели.

5. Упражнения на проекцию вектора на линейное подпространство

Матрицей шляпницей будем называть матрицу, которая выполняет следующую функцию

$$Hy = \hat{y}$$

В задаче регрессии эта матрица является выражается через X

$$Hy = \hat{y} = X\hat{\beta} = X(X^Y X)^{-1} X^T y$$

$$H = X(X^Y X)^{-1} X^T$$

Чему равны следующие выражения:

1. H^2 = ?

2. H^{2018} = ?

3. $(I - H)^2$ = ?

Решение: 1-й пример это двойная проекция на $Lin(X)$, поэтому $H^2 = H$, так как проекция уже спроецированного вектора в $Lin(X)$ совпадает с этим вектором. 2-й пример решатся ровно так же $H^{2018} = H$. В 3-м примере матрица $(I-H)$ является проекцией на ортогональное дополнение $Lin(X)$, т.е. проекция является вектором ошибок. Поэтому $(I - H)^2$ является двойной проекцией на ортогональное дополнение $Lin(X)$, т.е. $(I - H)^2 = (I - H)$

6. Геометрическая интерпретация транспонирования

Операцию транспонирования можно определить при помощи скалярного произведения не прибегая к фразе ”переставление столбцов и строк местами”.

Пусть есть 2 вектора x и y. Предположим над вектором x совершили преобразование A. Какое нужно совершить преобразование B над вектором y для того, чтобы выполнилось равенство:

$$< Ax, y >=< x, By >$$

Представим скалярное произведение в алгебраическом виде:

$$(Ax)^T y = x^T B y$$

$$x^T A^T y = x^T B y$$

Отсюда очевидно, что $B = A^T$.