Модуль 14.2. Методы приближения функций и обработки данных (часть III)

Метод наименьших квадратов (для обработки данных)

Задача о построении МНК-полинома

Предположим, что некоторые процессы или объекты описываются вещественными переменными X и Y, и по результатам наблюдений нужно выявить **функциональную** связь этих переменных.

Обычно одну из переменных рассматривают как причину (**фактор**, или объясняющую переменную), другую – как следствие (**отклик**, или объясняемую переменную).

Пусть X – фактор, Y – отклик. Рассмотрим метод построения функциональной зависимости Y от X в виде полинома степени не выше K.

Результаты наблюдений над процессами (объектами) запишем в виде пар значений

$$(X_i, Y_i),\ i = 1,\ldots n \tag{14.1}$$

где n есть количество наблюдений, i – номер наблюдения.

Точки с координатами (14.1) отметим на плоскости $X - Y$.

Если количество точек (n) больше, чем количество неизвестных коэффициентов полинома ($K+1$) , полином нужной степени может через эти точки не пройти.

Метод наименьших квадратов (МНК) позволяет строить полином нужной степени, который не всегда пройдет через заданные точки, но приблизится к ним оптимальным образом.

Чтобы сформулировать критерий оптимальности, нужно отличать значения отклика, полученные при сборе данных, от значений, вычисленных с помощью полинома.

Определение 1. **Истинными значениями отклика** Y называют значения $Y = Y_i,\ i = 1,\ldots n$, которые измерены при $X = X_i,\ i = 1,\ldots n$ и указаны в наборе данных (14.1).

Значения Y, вычисленные с помощью полинома, обозначим через $\hat{Y}$.
Искомый полином запишем в виде

$$\hat{Y} = b_0 + b_1 X + \ldots + b_K X^K \tag{14.2}$$

Определение 2. Значения отклика, вычисленные при $X = X_i,\ i = 1,\ldots n$, обозначим через $\hat{Y}_i,\ \ i = 1,\ldots n$:

$$\hat{Y}_i = b_0 + b_1 X_i + \ldots + b_K X_i^K,\ \ i = 1,\ldots n \tag{14.3}$$

Величины $\hat{Y}_i,\ \ i = 1,\ldots n$ **называют оценочными значениями отклика** Y.

Определение 3. **Остатками $\hat{\varepsilon}_i, \quad i = 1,...n$ называют разности истинных и оценочных значений отклика**

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i, \quad i = 1,...n \qquad (14.4)$$

Определение 4. Согласно методу наименьших квадратов (МНК), среди всех полиномов вида (14.2) **наилучшим** считается тот, которому соответствует **минимальная сумма квадратов остатков**.

Такой полином называют **МНК-полиномом**.

Пример 1

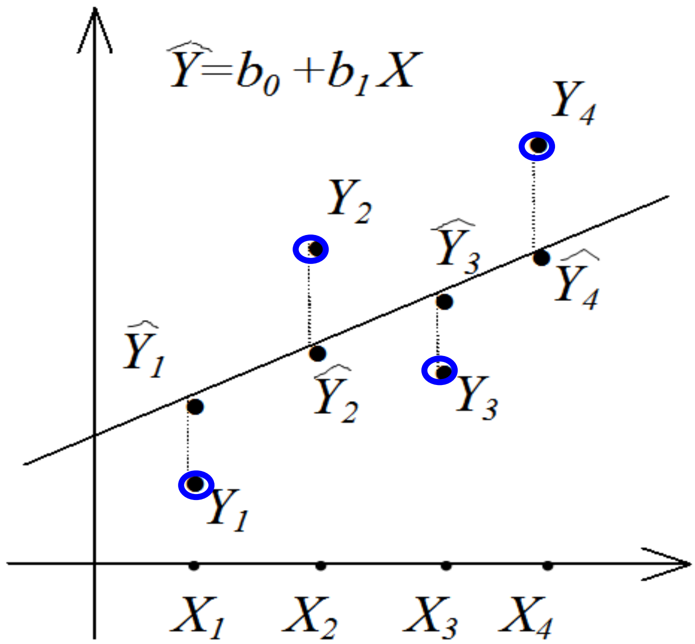


Рисунок 1

На рисунке показаны точками результаты 4-х наблюдений: $(X_i, Y_i), \ i = 1,..4,$ синий маркер. Через указанные точки нельзя провести полином степени $K = 1$

(нельзя провести прямую).

Поэтому показна МНК-прямая $\hat{Y} = b_0 + b_1 X$ и оценочные значения $\hat{Y}_i, \quad i = 1,...4$ (они расположены на МНК-прямой).

Пунктиром показаны абсолютные значения остатков $\hat{\varepsilon}_i, \quad i = 1,...4$.
Остатки с номерами 1, 3 отрицательны, с номерами 2, 4 – положительны.

МНК-прямая, показанная на рисунке, обеспечивает минимальную сумму квадратов остатков:

значение $S = \sum_{i=1}^{4} [\hat{\varepsilon}_i]^2$ для данной прямой минимально.

Способ построения МНК-полинома

Сумму квадратов остатков обозначим S и выясним, чему она равна:

$$S=\sum_{i=1}^{n}[\hat{\varepsilon}_i]^2=\sum_{i=1}^{n}[Y_i-\hat{Y}_i]^2=\sum_{i=1}^{n}[Y_i-(b_0+b_1X_i+...+b_KX_i^K)]^2 \qquad (14.5)$$

Для построения МНК-полинома необходимо найти такие коэффициенты $b_j,\ j=0,...K$, для которых S принимает минимальное значение.

То есть нужно решить задачу

$$S(b_0,b_1,...b_K)=\sum_{i=1}^{n}[Y_i-(b_0+b_1X_i+...+b_KX_i^K)]^2\to\min \qquad (14.6)$$

при $(b_0,b_1,...b_K)\in R^{K+1}$

При решении задачи оптимизации (14.6) S рассматривается как функция $K+1$ переменной $b_j,\ j=0,...K$, а значения $X_i,Y_i,\ i=1,...n$ есть числа: они уже получены в результате замеров (14.1).

Чтобы сформулировать результат о существовании, единственности и способе отыскания МНК-полинома, введем дополнительные обозначения.

Значения фактора X и функции фактора X запишем в матрицу $\boldsymbol{X}$, которую называют **матрицей регрессоров**.

Матрица $\boldsymbol{X}$ имеет размерность $n\times(K+1)$. Ее строки соответствуют наблюдениям, а столбцы – регрессорам.

$$\boldsymbol{X}=\begin{bmatrix}1 & X_1 & X_1^2 & ... & X_1^K\\ 1 & X_2 & X_2^2 & ... & X_2^K\\ ... & ... & ... & ... & ...\\ 1 & X_n & X_n^2 & ... & X_n^K\end{bmatrix} \qquad (14.7)$$

Первый столбец матрицы $\boldsymbol{X}$ состоит из единиц. Во втором столбце матрицы $\boldsymbol{X}$ указаны значения фактора X по всем замерам: $X_i,\ i=1,...n$. В третьем и следующих столбцах – результат возведения фактора X в соответствующую степень – от 2 до K:

$X_i^2,\ i=1,...n,$

… … …

$X_i^K,\ i=1,...n$

Дополнительно к (14.7) нужны **обозначения для векторов**: истинные значения отклика Y запишем как вектор $\overline{\boldsymbol{Y}}$ размерности n; столбцы матрицы $\boldsymbol{X}$ – как векторы $\overline{\boldsymbol{X}}^{(j)},\ j=0,..K$ размерности n:

$$\overline{\boldsymbol{Y}}=\begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix},\quad \overline{\boldsymbol{X}}^{(0)}=\begin{bmatrix} 1 \\ 1 \\ \dots \\ 1 \end{bmatrix},\ \overline{\boldsymbol{X}}^{(1)}=\begin{bmatrix} X_1 \\ X_2 \\ \dots \\ X_n \end{bmatrix},\quad \dots\dots\dots \quad \overline{\boldsymbol{X}}^{(K)}=\begin{bmatrix} X_1^K \\ X_2^K \\ \dots \\ X_n^K \end{bmatrix}$$

Утверждение 1. Для любого набора данных (14.1), такого, что ранг матрицы $\boldsymbol{X}$ равен $K+1$, МНК-полином (14.2) **существует** и **является единственным**, а его коэффициенты $b_j,\ j=1,...K$ **являются решением нормальной системы уравнений**:

$$\frac{\partial S}{\partial b_j}=0,\ j=0,...K \tag{14.8}$$

Система (14.8) представляет собой СЛАУ с неизвестными $b_j,\ j=1,...K$:

$$\begin{bmatrix} (\overline{X}^{(0)},\overline{X}^{(0)}) & (\overline{X}^{(0)},\overline{X}^{(1)}) & \dots & (\overline{X}^{(0)},\overline{X}^{(K)}) \\ (\overline{X}^{(1)},\overline{X}^{(0)}) & (\overline{X}^{(1)},\overline{X}^{(1)}) & \dots & (\overline{X}^{(1)},\overline{X}^{(K)}) \\ \dots & \dots & \dots & \dots \\ (\overline{X}^{(K)},\overline{X}^{(0)}) & (\overline{X}^{(K)},\overline{X}^{(1)}) & \dots & (\overline{X}^{(K)},\overline{X}^{(K)}) \end{bmatrix}\cdot\begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_K \end{bmatrix}=\begin{bmatrix} (\overline{Y},\overline{X}^{(0)}) \\ (\overline{Y},\overline{X}^{(1)}) \\ \dots \\ (\overline{Y},\overline{X}^{(K)}) \end{bmatrix} \tag{14.9}$$

Символом $(*\cdot,*)$ обозначено скалярное произведение в пространстве R^n.

Доказательство

Рассмотрим задачу оптимизации (14.6):

$$S(b_0,b_1,...b_K)\underset{b\in R^{K+1}}{\to}\min.$$

Точки, подозрительные на экстремум, находим из условий

$$\frac{\partial S}{\partial b_j}=0,\ j=0,...K$$

Шаг за шагом покажем:

1) линейная независимость столбцов матрицы $\boldsymbol{X}$ обеспечивает существование и единственность решения СЛАУ (14.9);

2) в силу линейной независимости указанных столбцов единственное решение СЛАУ (14.9) является точкой локального минимума;

3) в силу свойств функционала $S(b_0,b_1,...b_K)$ локальный минимум является глобальным, и это означает, что задача оптимизации (14.6) решена.

Шаг I

Покажем, что систему (14.8) можно записывать в виде (14.9).

Используя (14.5), запишем частные производные функционала $S(b_0, b_1, \dots b_K)$:

$$\begin{cases} \dfrac{\partial S}{\partial b_0} = -2\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + \dots + b_K X_i^K)) \\ \dfrac{\partial S}{\partial b_1} = -2\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + \dots + b_K X_i^K))\ X_i \\ \dots \\ \dfrac{\partial S}{\partial b_K} = -2\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + \dots + b_K X_i^K))\ X_i^K \end{cases} \tag{14.10}$$

В соответствии уравнениями (14.8) приравниваем каждую из частных производных к нулю, полученные выражения делим на 2:

$$\begin{cases} \dfrac{1}{2}\cdot\dfrac{\partial S}{\partial b_0} = -\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + \dots + b_K X_i^K)) = 0 \\ \dfrac{1}{2}\cdot\dfrac{\partial S}{\partial b_1} = -\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + \dots + b_K X_i^K))\ X_i = 0 \\ \dots \\ \dfrac{1}{2}\cdot\dfrac{\partial S}{\partial b_K} = -\sum\limits_{i=1}^{n}(Y_i - (b_0 + b_1 X_i + \dots + b_K X_i^K))\ X_i^K = 0 \end{cases} \tag{14.11}$$

Слагаемые, не зависящие от коэффициентов $b_j, j = 0, \dots K$, переносим в правую часть каждого уравнения:

$$\begin{cases} \sum\limits_{i=1}^{n}(b_0 + b_1 X_i + \dots + b_K X_i^K) = \sum\limits_{i=1}^{n} Y_i \\ \sum\limits_{i=1}^{n}(b_0 + b_1 X_i + \dots + b_K X_i^K)\ X_i = \sum\limits_{i=1}^{n} Y_i X_i \\ \dots \\ \sum\limits_{i=1}^{n}(b_0 + b_1 X_i + \dots + b_K X_i^K)\ X_i^K = \sum\limits_{i=1}^{n} Y_i X_i^K \end{cases} \tag{14.12}$$

Таким образом, на основе (14.8) получена СЛАУ с неизвестными $b_j, j = 0, \dots K$.

Запишем СЛАУ в векторном виде:

$$
\begin{bmatrix}
n & \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 & \dots & \sum_{i=1}^{n} X_i^K \\
\sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 & \sum_{i=1}^{n} X_i^3 & \dots & \sum_{i=1}^{n} X_i^{K+1} \\
\dots & \dots & \dots & \dots & \dots \\
\sum_{i=1}^{n} X_i^K & \sum_{i=1}^{n} X_i^{K+1} & \sum_{i=1}^{n} X_i^{K+2} & \dots & \sum_{i=1}^{n} X_i^{2K}
\end{bmatrix}
\times
\begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_K \end{bmatrix}
=
\begin{bmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} Y_i X_i \\ \dots \\ \sum_{i=1}^{n} Y_i X_i^K \end{bmatrix}
$$

Несложно проверить, что элемент матрицы СЛАУ, расположенный в строке с номером l и столбце с номером m, имеет вид

$$
\sum_{i=1}^{n} X_i^l \cdot X_i^m = \sum_{i=1}^{n} X_i^{l+m} \tag{14.13}
$$

Данный элемент совпадает со скалярным произведением столбцов матрицы $\boldsymbol{X}$ с номерами l и m:

$$
(\overline{X}^{(l)}, \overline{X}^{(m)}) = \sum_{i=1}^{n} X_i^{l+m} \tag{14.14}
$$

Аналогично для вектора, указанного в правой части СЛАУ: элемент, расположенный в строке с номером l имеет вид

$$
\sum_{i=1}^{n} Y_i \ X_i^l \tag{14.15}
$$

Он совпадает со скалярным произведением вектора $\overline{\boldsymbol{Y}}$ и столбца матрицы $\boldsymbol{X}$ с номером l:

$$
(\overline{Y}, \overline{X}^{(l)}) = \sum_{i=1}^{n} Y_i \ X_i^l \tag{14.16}
$$

Учитывая соответствие (14.13)-(14.16), используя ранее введенные обозначения для столбцов матрицы $\boldsymbol{X}$, запишем СЛАУ (14.12) в виде (14.9)

$$
\begin{bmatrix}
(\overline{X}^{(0)}, \overline{X}^{(0)}) & (\overline{X}^{(0)}, \overline{X}^{(1)}) & \dots & (\overline{X}^{(0)}, \overline{X}^{(K)}) \\
(\overline{X}^{(1)}, \overline{X}^{(0)}) & (\overline{X}^{(1)}, \overline{X}^{(1)}) & \dots & (\overline{X}^{(1)}, \overline{X}^{(K)}) \\
\dots & \dots & \dots & \dots \\
(\overline{X}^{(K)}, \overline{X}^{(0)}) & (\overline{X}^{(K)}, \overline{X}^{(1)}) & \dots & (\overline{X}^{(K)}, \overline{X}^{(K)})
\end{bmatrix}
\cdot
\begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_K \end{bmatrix}
=
\begin{bmatrix} (\overline{Y}, \overline{X}^{(0)}) \\ (\overline{Y}, \overline{X}^{(1)}) \\ \dots \\ (\overline{Y}, \overline{X}^{(K)}) \end{bmatrix}
$$

Таким образом, представление нормальной системы уравнений (14.8) в виде СЛАУ (14.9) доказано.

Ш а г II

Исследуем возможность решения СЛАУ.

Матрица СЛАУ (14.9) является матрицей Грама для векторов, являющихся столбцами матрицы $\boldsymbol{X}$:

$$Gr(\overline{X}^{(0)},...\overline{X}^{(K)})=\begin{bmatrix} (\overline{X}^{(0)},\overline{X}^{(0)}) & (\overline{X}^{(0)},\overline{X}^{(1)}) & ... & (\overline{X}^{(0)},\overline{X}^{(K)}) \\ (\overline{X}^{(1)},\overline{X}^{(0)}) & (\overline{X}^{(1)},\overline{X}^{(1)}) & ... & (\overline{X}^{(1)},\overline{X}^{(K)}) \\ ... & ... & ... & ... \\ (\overline{X}^{(K)},\overline{X}^{(0)}) & (\overline{X}^{(K)},\overline{X}^{(1)}) & ... & (\overline{X}^{(K)},\overline{X}^{(K)}) \end{bmatrix}$$

По условию Утверждения 1, ранг $\boldsymbol{X}$ равен $K+1$. Это означает, что столбцы матрицы $\boldsymbol{X}$ линейно независимы и поэтому матрица Грама не вырождена и положительно определена.

$$\det\ Gr\ (\overline{X}^{(0)},\ \overline{X}^{(1)}...\overline{X}^{(K)})\neq 0 \tag{14.17}$$

$$Gr\ (\overline{X}^{(0)},\ \overline{X}^{(1)}...\overline{X}^{(K)})>0 \tag{14.18}$$

Те же самые свойства имеет матрица СЛАУ.

Поэтому при любой правой части СЛАУ (14.9) ее решение существует и единственно, а функционал $S(b_0,b_1,...b_K)$ имеет единственную точку, подозрительную на экстремум.

Ш а г III

Исследуем критическую точку с помощью матрицы вторых производных.

$$S''(b_0,b_1...b_K)=\begin{bmatrix} \dfrac{\partial^2 S}{\partial b_0^2} & ... & \dfrac{\partial^2 S}{\partial b_0\,\partial b_K} \\ & ... & \\ ... & ... & ... \\ \dfrac{\partial^2 S}{\partial b_K\,\partial b_0} & ... & \dfrac{\partial^2 S}{\partial b_K^2} \end{bmatrix} \tag{14.19}$$

Вычисляя вторые частные производные, убеждаемся в том,

что для функционала $S(b_0,b_1,...b_K)$

матрица $S''(b_0,b_1,...b_K)$

с точностью до множителя 2

совпадает с матрицей Грама

линейно независимых столбцов матрицы $\boldsymbol{X}$:

$$S''(b_0,\ b_1...b_K)=2\cdot Gr(\overline{X}^{(0)},\overline{X}^{(1)}...\overline{X}^{(K)}) \tag{14.20}$$

Приведем некоторые примеры совпадения их элементов:

$$\begin{cases} \dfrac{\partial^2 S}{\partial b_0^2} = 2\,n \qquad \dfrac{\partial^2 S}{\partial b_0 \partial b_1} = 2\sum_{i=1}^{n} X_i \quad \ldots\ldots\ldots \quad \dfrac{\partial^2 S}{\partial b_0 \partial b_K} = 2\sum_{i=1}^{n} X_i^K \\ \ldots \\ \ldots \qquad\qquad \ldots \qquad\qquad \ldots\ldots\ldots \quad \dfrac{\partial^2 S}{\partial b_K^2} = 2\sum_{i=1}^{n} X_i^{2K} \end{cases}$$

Из (14.20) следует, что $S''(b_0, b_1 \ldots b_K)$, во-первых, не зависит от коэффициентов $b_j, j = 0,\ldots K$; во-вторых, положительно определена:

$$S''(b_0, b_1 \ldots b_K) = S'' > 0 \qquad (14.21)$$

Таким образом, точка, подозрительная на экстремум, является точкой локального минимума.

Решение СЛАУ (14.9) является точкой локального минимума функционала $S(b_0, b_1, \ldots b_K)$.

Аналогично Утверждению 4 Модуля 14.1 доказывается:

Решение нормальной системы уравнений (14.9), являясь точкой локального минимума функционала $S(b_0, b_1, \ldots b_K)$, **является решением задачи минимизации (14.6), то есть глобальным минимумом** $S(b_0, b_1, \ldots b_K)$.

Считаем, что Утверждение 1 доказано.

Следствие. Если в наборе данных (14.1) истинные значения отклика Y представляют собой значения некоторого полинома степени не выше K, а именно

$$Y_i = a_0 + a_1 X_i + \ldots + a_K X_i^K, \quad i = 1,\ldots n \qquad (14.22)$$

методом наименьших квадратов будет построен именно этот полином:

$$\hat{Y} = b_0 + b_1 X + \ldots + b_K X^K,$$

где $b_j = a_j,\ j = 0,\ldots K$

Доказательство

Если в наборе данных (14.1) истинные значения отклика Y представляют собой значения полинома (14.22) степени не выше K,

для данного полинома истинные и оценочные значения отклика совпадают.

Тогда остатки равны нулю

$\hat{\varepsilon}_i = Y_i - \hat{Y}_i = Y_i - Y_i = 0, \quad i = 1,...n..$

и полиному (14.22) соответствует нулевая сумма квадратов остатков.

Значение функционала $S(b_0, b_1,...b_K)$ меньше нуля не бывает.

Единственным решением (14.6) будет именно (14.22).

Пример 2

В таблице приведены результаты 4-х замеров:

X_i	0	1	3	6
Y_i	0.1	0.8	3.05	5.95

Нужно построить линейную зависимость Y от X методом наименьших квадратов и затем приблизить данные полиномом 2-й степени. В каждом из случаев указать, какой функционал должен быть минимизирован.

Решение

1) Строим полином степени $K = 1$, то есть МНК-прямую.

Количество наблюдений $n = 4$. Полином запишем в виде $\hat{Y} = b_0 + b_1 X$.

Параметры полинома должны быть решением задачи оптимизации

$$S(b_0, b_1) = \sum_{i=1}^{4} [\, Y_i - (b_0 + b_1 X_i] \;^2 \to \min$$

В данной задаче

$$S(b_0, b_1) = [0.1 - (b_0 + b_1 \cdot 0)]^2 + [0.8 - (b_0 + b_1 \cdot 1)]^2 +$$
$$+ [3.05 - (b_0 + b_1 \cdot 3)]^2 + [5.95 - (b_0 + b_1 \cdot 6)]^2$$

Если формулы (14.9) были «забыты», решение задачи можно найти, записывая и решая нормальную систему уравнений

$$\frac{\partial S}{\partial b_0} = 0, \quad \frac{\partial S}{\partial b_1} = 0.$$

В этом примере о способе записи (14.9) не забываем.

Вектор истинных значений отклика $\overline{\boldsymbol{Y}}$ и матрица регрессоров $\boldsymbol{X}$ размерности 4×2 равны

$$\overline{\boldsymbol{Y}} = \begin{bmatrix} 0.1 \\ 0.8 \\ 3.05 \\ 5.95 \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 3 \\ 1 & 6 \end{bmatrix}$$

Столбцы матрицы регрессоров определяют векторы

$$\overline{X}^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \overline{X}^{(1)} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \end{bmatrix}$$

Нормальная система уравнений для отыскания b_j, $j = 0,1$ имеет вид:

$$\begin{bmatrix} 1+1+1+1 & 1\cdot 0+1\cdot 1+3\cdot 1+6\cdot 1 \\ 1\cdot 0+1\cdot 1+1\cdot 3+1\cdot 6 & 0\cdot 0+1\cdot 1+3\cdot 3+6\cdot 6 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} =$$

$$= \begin{bmatrix} 0.1\cdot 1+0.8\cdot 1+3.05\cdot 1+5.95\cdot 1 \\ 0.1\cdot 0+0.8\cdot 1+3.05\cdot 3+5.95\cdot 6 \end{bmatrix}$$

то есть

$$\begin{bmatrix} 4 & 10 \\ 10 & 46 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} 9.9 \\ 45.65 \end{bmatrix}$$

Решение СЛАУ и МНК-прямая показаны на Рисунке 2.

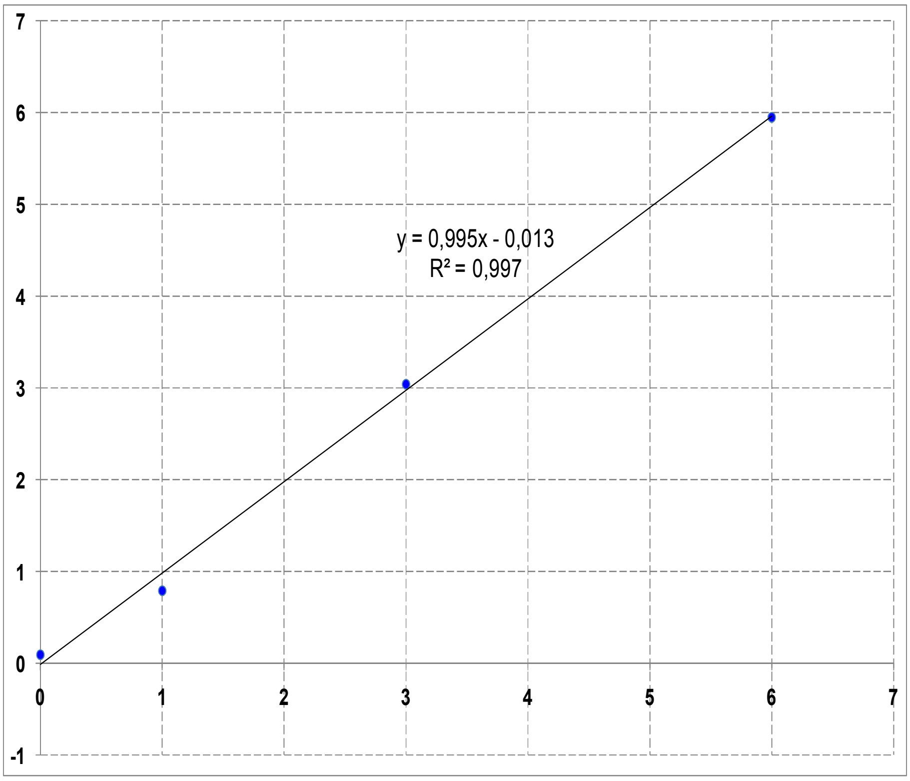


Рисунок 2

2) Строим полином степени $K = 2$, то есть МНК-параболу.

Количество наблюдений $n = 4$. Полином запишем в виде

$\hat{Y} = b_0 + b_1 X + b_2 X^2$.

Параметры полинома должны быть решением задачи оптимизации

$$S(b_0, b_1, b_2) = \sum_{i=1}^{4} [\, Y_i - (b_0 + b_1 X_i + b_2 X_i^2)]^{\,2} \to \min$$

В данной задаче

$$S(b_0, b_1, b_2) = [0.1 - b_0]^2 + [0.8 - (b_0 + b_1 + b_2)]^2 +$$
$$+ [3.05 - (b_0 + b_1 \cdot 3 + b_2 \cdot 9)]^2 + [5.95 - (b_0 + b_1 \cdot 6 + b_2 \cdot 36)]^2$$

Как и в предыдущем случае, если формулы (14.9) «забыты», решение задачи можно найти, записывая и решая нормальную систему уравнений

$$\frac{\partial S}{\partial b_0} = 0, \quad \frac{\partial S}{\partial b_1} = 0, \quad \frac{\partial S}{\partial b_2} = 0.$$

Как и в предыдущем случае, о формулах (14.9) не забываем.

Вектор истинных значений отклика $\overline{\boldsymbol{Y}}$ и матрица регрессоров $\boldsymbol{X}$ размерности 4×3 равны

$$\overline{\boldsymbol{Y}} = \begin{bmatrix} 0.1 \\ 0.8 \\ 3.05 \\ 5.95 \end{bmatrix} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 3 & 9 \\ 1 & 6 & 36 \end{bmatrix}$$

Столбцы матрицы регрессоров определяют векторы

$$\overline{\boldsymbol{X}}^{(0)} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad \overline{\boldsymbol{X}}^{(1)} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \end{bmatrix} \qquad \overline{\boldsymbol{X}}^{(2)} = \begin{bmatrix} 0 \\ 1 \\ 9 \\ 36 \end{bmatrix}$$

Нормальная система уравнений для отыскания $b_j,\ j = 0, 1, 2$ имеет вид:

$$\begin{bmatrix} 4 & 10 & 46 \\ 10 & 46 & 244 \\ 46 & 244 & 1378 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ b_2 \end{bmatrix} = \begin{bmatrix} 9.9 \\ 45.65 \\ 242.45 \end{bmatrix}$$

Решение СЛАУ и МНК-парабола показаны на Рисунке 3.

3) Докажите, что попытка построить МНК полином степени $K = 3$

$$\hat{Y} = b_0 + b_1 X + b_2 X^2 + b_3 X^3$$

приведет к тому, что в результате минимизации функционала

$$S(b_0, b_1, b_2, b_3) = \sum_{i=1}^{4} [\, Y_i - (b_0 + b_1 X_i + b_2 X_i^2 + b_3 X_i^3)]^{\,2} \to \min$$

будет построен **интерполяционный полином** степени 3, см. Рисунок 4.

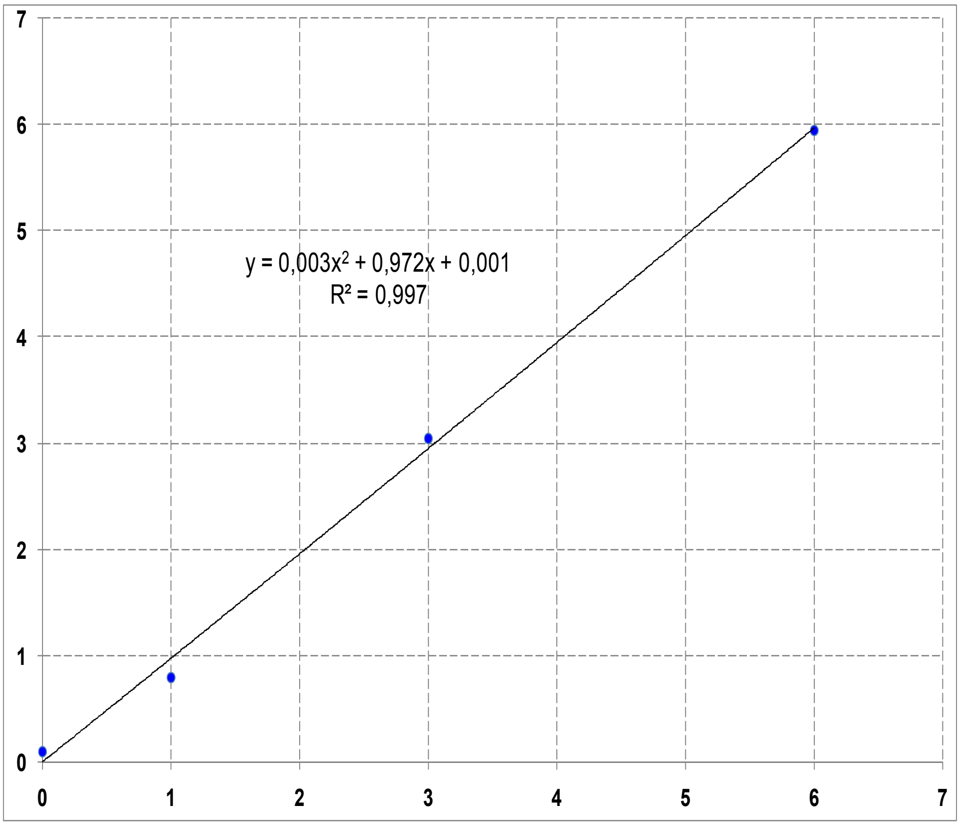


Рисунок 3

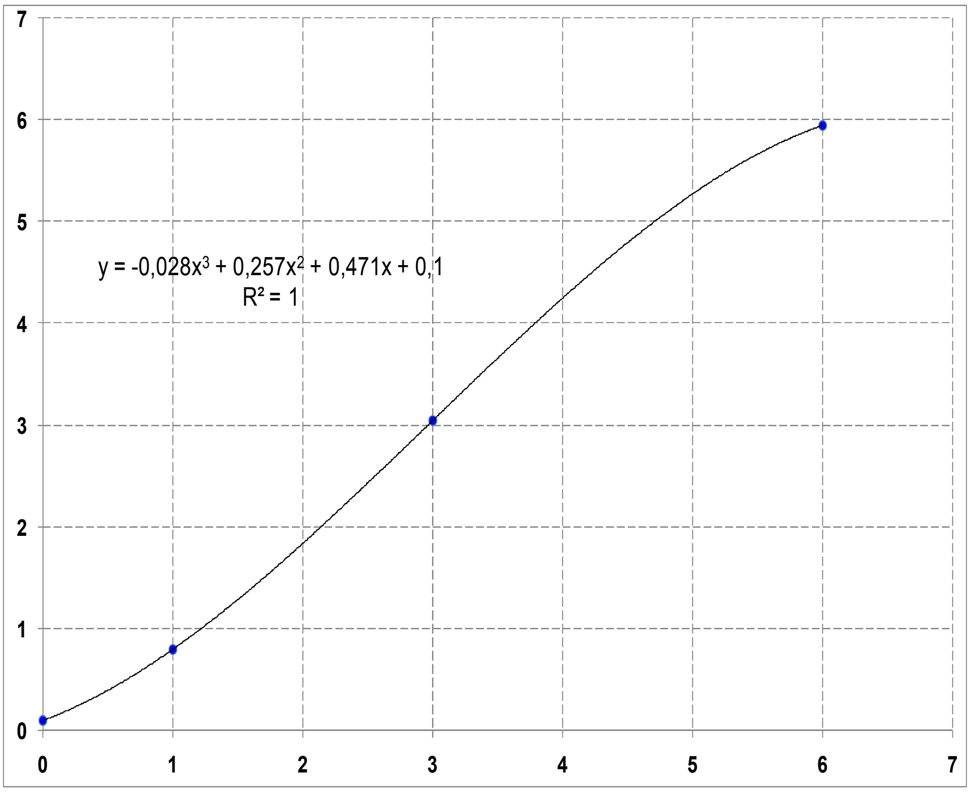


Рисунок 4

Задача о построении обобщенного МНК-полинома

Рассмотрим обобщение метода наименьших квадратов
на случай нескольких объясняющих переменных (нескольких факторов) и (или)
случай более сложных функциональных зависимостей
между откликом и факторами.

В этом разделе важно, чтобы функциональная зависимость отклика от факторов оставалась линейной относительно неизвестных параметров.

Пусть количество объясняющих переменных (факторов) равно m.

Фактор с номером j обозначим $X^{(j)}$, $j=1,\ldots m$.

Через X обозначим весь набор факторов:

$X=(X^{(1)},X^{(2)},\ldots X^{(m)})$, $X\in R^m$.

Отклик обозначим Y .

Результаты наблюдений над процессами (объектами) запишем в виде наборов значений

$$X_i=(X_i^{(1)},X_i^{(2)},\ldots X_i^{(m)}),\ Y_i,\ i=1,\ldots n \qquad (14.23)$$

где n есть количество наблюдений, i – номер наблюдения,

$X_i\in R^m, i=1,\ldots n$ – набор значений факторов для наблюдения (объекта) с номером i,

$Y_i,\ i=1,\ldots n$ – значение отклика Y для наблюдения (объекта) с номером i.

Для описания функциональной зависимости отклика Y от факторов $X^{(j)}$, $j=1,\ldots m$ будут использованы функции

$\varphi_0(X),\ \varphi_1(X),\ldots \varphi_K(X),$

а сама зависимость найдена в виде линейной комбинации этих функций:

$$Y=b_0\cdot\varphi_0(X)+b_1\cdot\varphi_1(X)+\ldots+b_K\cdot\varphi_K(X) \qquad (14.24)$$

Выражение (14.24) называют обобщенным полиномом.

Примеры обобщенных полиномов

1) Если фактор только один, то есть $X\in R,\ m=1,$ и для описания зависимости Y от X выбраны функции

$\varphi_0(X)=1,\ \varphi_1(X)=X,\ldots \varphi_K(X)=X^K$

обобщенный полином (14.24) представляет собой полином степени K:

$Y=b_0+b_1X+\ldots+b_KX^K$

2) Если $X \in R,\ m=1,$ и для описания зависимости Y от X выбраны функции

$$\varphi_0(X)=X,\ \varphi_1(X)=X^2$$

обобщенный полином является полиномом степени 2 относительно X без константы:

$$Y=b_0X+b_1X^2$$

3) Если $X \in R,\ m=1,$ и для описания зависимости Y от X выбраны функции

$$\varphi_0(X)=X,\ \varphi_1(X)=e^X$$

обобщенный полином есть функция вида

$$Y=b_0X+b_1e^X$$

4) Если есть два фактора, то есть $X \in R^2,\ m=2,$ и выбраны функции

$$\varphi_0(X)=1,\ \varphi_1(X)=X^{(1)},\dots\varphi_2(X)=X^{(1)}\sin\left(X^{(2)}\right)$$

обобщенный полином есть функция вида

$$Y=b_0+b_1X^{(1)}+b_2X^{(1)}\sin(X^{(2)})$$

5) Если есть два фактора, то есть $X \in R^2,\ m=2,$ и выбраны функции

$$\varphi_0(X)=1,\ \varphi_1(X)=X^{(1)},\dots\varphi_2(X)=X^{(2)}$$

обобщенный полином есть линейная функция вида

$$Y=b_0+b_1X^{(1)}+b_2X^{(2)}$$

Функции $\varphi_0(X),\ \varphi_1(X),\dots\varphi_K(X)$ выбирают, руководствуясь характером функциональной зависимости отклика от факторов, а неизвестные параметры $b_j,\ j=0,\dots K$ можно подобрать методом наименьших квадратов (МНК).

При этом в классе функций вида (14.24) будет выбрана такая, которая приблизит результаты наблюдений (14.23) оптимальным образом.

Как и в предыдущем разделе, чтобы сформулировать критерий оптимальности, нужно отличать значения отклика, полученные при сборе данных, от вычисляемых значений.

Определение 5. **Истинными значениями отклика** Y называют значения

$$Y = Y_i, \quad i = 1,...n,$$

измеренные при $X = X_i, \quad i = 1,...n$ и указанные в наборе значений (14.23).

Значения Y, вычисляемые с помощью (14.24), обозначим через $\hat{Y}$.
Искомую зависимость запишем в виде

$$\hat{Y} = b_0 \cdot \varphi_0(X) + b_1 \cdot \varphi_1(X) + ... + b_K \cdot \varphi_K(X) \qquad (14.25)$$

Определение 6. Значения отклика, **вычисленные** при $X = X_i, \quad i = 1,...n$, обозначим через $\hat{Y}_i, \quad i = 1,...n$:

$$\hat{Y} = b_0 \cdot \varphi_0(X_i) + b_1 \cdot \varphi_1(X_i) + ... + b_K \cdot \varphi_K(X_i), \qquad i = 1,...n \qquad (14.26)$$

Величины $\hat{Y}_i, \quad i = 1,...n$ **называют оценочными значениями отклика** Y.

Определение 7. **Остатками** $\hat{\varepsilon}_i, \quad i = 1,...n$ **называют разности истинных и оценочных значений отклика**

$$\hat{\varepsilon}_i = Y_i - \hat{Y}_i, \quad i = 1,...n \qquad (14.27)$$

Определение 8. Согласно методу наименьших квадратов (МНК), среди всех обобщенных полиномов вида (14.25) **наилучшим** считается тот, которому соответствует **минимальная сумма квадратов остатков**.

Такой полином называют обобщенным **МНК-полиномом**.

Способ построения обобщенного МНК-полинома

Запишем, чему равна S **– сумма квадратов остатков:**

$$S = \sum_{i=1}^{n} [\hat{\varepsilon}_i]^2 = \sum_{i=1}^{n} [Y_i - \hat{Y}_i]^2 =$$

$$= \sum_{i=1}^{n} [Y_i - (b_0 \cdot \varphi_0(X_i) + b_1 \cdot \varphi_1(X_i) + ... + b_K \cdot \varphi_K(X_i))]^2$$

(14.28)

Для построения обобщенного МНК-полинома необходимо найти такие коэффициенты $b_j, \; j = 0,...K$, **для которых** S **принимает минимальное значение, то есть решить задачу**

$$S(b_0, b_1 ... b_K) = \sum_{i=1}^{n} [\, Y_i - (b_0 \cdot \varphi_0(X_i) + b_1 \cdot \varphi_1(X_i) + ... + b_K \cdot \varphi_K(X_i))]^{\,2} \to \min$$

(14.29)

при $(\, b_0, b_1, ... b_K) \in R^{K+1}$

Здесь S рассматривается как функция $K+1$ переменной $b_j,\ j=0,...K$, а значения

$X_i = (X_i^{(1)}, X_i^{(2)}, ... X_i^{(m)}),\ Y_i,\ i=1,...n$

указаны в наборе данных (14.23) и являются числами.

Как и в случае «обычных» полиномов, введем дополнительные обозначения.

Значения факторов X и функции $\varphi_0(X),\ \varphi_1(X), ... \varphi_K(X)$ запишем в матрицу $\boldsymbol{\Phi}$, которую называют **матрицей регрессоров**.

Матрица $\boldsymbol{\Phi}$ имеет размерность $n \times (K+1)$. Ее строки соответствуют наблюдениям, а столбцы – регрессорам.

$$\boldsymbol{\Phi} = \begin{bmatrix} \varphi_0(X_1) & \varphi_1(X_1) & \varphi_2(X_1) & ... & \varphi_K(X_1) \\ \varphi_0(X_2) & \varphi_1(X_2) & \varphi_2(X_2) & ... & \varphi_K(X_2) \\ ... & ... & ... & ... & ... \\ \varphi_0(X_n) & \varphi_1(X_n) & \varphi_2(X_n) & ... & \varphi_K(X_n) \end{bmatrix} \qquad (14.30)$$

Первый столбец матрицы $\boldsymbol{\Phi}$ состоит из значений функции $\varphi_0(X)$ по всем наблюдениям $i=1,...n,$ то есть по всем точкам $X_i = (X_i^{(1)}, X_i^{(2)}, ... X_i^{(m)}),\ i=1,...n$.

Во втором столбце матрицы $\boldsymbol{\Phi}$ указаны значения функции $\varphi_1(X)$ по всем наблюдениям $i=1,...n$.

В третьем и следующих столбцах – значения функций $\varphi_j(X),\ \ j=2,...K$ по всем наблюдениям $i=1,...n$.

Дополнительно к (14.30) нужны **обозначения для векторов**: истинные значения отклика Y запишем как вектор $\overline{\boldsymbol{Y}}$ размерности n; столбцы матрицы $\boldsymbol{\Phi}$ – как векторы $\overline{\boldsymbol{\Phi}}^{(j)}, j=0,..K$ размерности n:

$$\overline{\boldsymbol{Y}} = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_n \end{bmatrix}, \quad \overline{\boldsymbol{\Phi}}^{(0)} = \begin{bmatrix} \varphi_0(X_1) \\ \varphi_0(X_2) \\ \dots \\ \varphi_0(X_n) \end{bmatrix}, \quad \overline{\boldsymbol{\Phi}}^{(1)} = \begin{bmatrix} \varphi_1(X_1) \\ \varphi_1(X_2) \\ \dots \\ \varphi_2(X_n) \end{bmatrix}, \quad \dots\dots\dots \quad \overline{\boldsymbol{\Phi}}^{(K)} = \begin{bmatrix} \varphi_K(X_1) \\ \varphi_K(X_2) \\ \dots \\ \varphi_K(X_n) \end{bmatrix}$$

Утверждение 2. Для любого набора данных (14.23), такого, что ранг матрицы $\boldsymbol{\Phi}$ равен $K+1$, обощенный МНК-полином (14.24) **существует** и **является единственным**, а его коэффициенты $b_j, j=1,...K$ **являются решением нормальной системы уравнений**:

$$\frac{\partial S}{\partial b_j} = 0, \ j = 0,...K \qquad (14.31)$$

Система (14.31) представляет собой СЛАУ с неизвестными $b_j, j=1,...K$:

$$\begin{bmatrix} (\overline{\boldsymbol{\Phi}}^{(0)},\overline{\boldsymbol{\Phi}}^{(0)}) & (\overline{\boldsymbol{\Phi}}^{(0)},\overline{\boldsymbol{\Phi}}^{(1)}) & \dots & (\overline{\boldsymbol{\Phi}}^{(0)},\overline{\boldsymbol{\Phi}}^{(K)}) \\ (\overline{\boldsymbol{\Phi}}^{(1)},\overline{\boldsymbol{\Phi}}^{(0)}) & (\overline{\boldsymbol{\Phi}}^{(1)},\overline{\boldsymbol{\Phi}}^{(1)}) & \dots & (\overline{\boldsymbol{\Phi}}^{(1)},\overline{\boldsymbol{\Phi}}^{(K)}) \\ \dots & \dots & \dots & \dots \\ (\overline{\boldsymbol{\Phi}}^{(K)},\overline{\boldsymbol{\Phi}}^{(0)}) & (\overline{\boldsymbol{\Phi}}^{(K)},\overline{\boldsymbol{\Phi}}^{(1)}) & \dots & (\overline{\boldsymbol{\Phi}}^{(K)},\overline{\boldsymbol{\Phi}}^{(K)}) \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_K \end{bmatrix} = \begin{bmatrix} (\overline{\boldsymbol{Y}},\overline{\boldsymbol{\Phi}}^{(0)}) \\ (\overline{\boldsymbol{Y}},\overline{\boldsymbol{\Phi}}^{(1)}) \\ \dots \\ (\overline{\boldsymbol{Y}},\overline{\boldsymbol{\Phi}}^{(K)}) \end{bmatrix} \qquad (14.32)$$

Символом $(*\cdot,*)$ обозначено скалярное произведение в пространстве R^n.

Утверждение 2 доказывается аналогично Утверждению 1.

Пример 3

В таблице приведены результаты 4-х замеров:

X_i	0	1	3	6
Y_i	0.1	0.8	3.05	5.95

Нужно построить линейную зависимость Y от X (без константы) методом наименьших квадратов. Указать, какой функционал будет минимизирован.

Решение

1) Количество наблюдений $n=4$. Обобщенный полином запишем в виде $\hat{Y} = b_0 X$. Считаем, что $\varphi_0(X) = X$ и $K=0$.

Параметр полинома должен быть решением задачи оптимизации

$$S(b_0) = \sum_{i=1}^{4} [\, Y_i - b_0 X_i \,]^{\,2} \to \min$$

В данной задаче

$$S(b_0) = [0.1 - b_0 \cdot 0]^2 + [0.8 - b_0 \cdot 1]^2 +$$

$$+ [3.05 - b_0 \cdot 3]^2 + [5.95 - b_0 \cdot 6)]^2$$

Если формулы (14.32) были «забыты», решение задачи можно найти, записывая и решая нормальную систему уравнений. В данном случае она состоит из одного уравнения

$$\frac{\partial S}{\partial b_0} = 0.$$

Все-таки о способе записи (14.32) удобнее не забывать.

Вектор истинных значений отклика $\overline{\boldsymbol{Y}}$ и матрица регрессоров $\boldsymbol{\Phi}$ размерности 4×1 равны

$$\overline{\boldsymbol{Y}} = \begin{bmatrix} 0.1 \\ 0.8 \\ 3.05 \\ 5.95 \end{bmatrix} \qquad \boldsymbol{\Phi} = \begin{bmatrix} \varphi_0(0) \\ \varphi_0(1) \\ \varphi_0(3) \\ \varphi_0(6) \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \end{bmatrix}$$

Единственный столбец матрицы регрессоров соответствует вектору

$$\overline{\boldsymbol{\Phi}}^{(0)} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \end{bmatrix}$$

Нормальная система уравнений для отыскания b_0 имеет вид

$$(\overline{\boldsymbol{\Phi}}^{(0)}, \overline{\boldsymbol{\Phi}}^{(0)}) \cdot b_0 = (\overline{\boldsymbol{Y}}, \overline{\boldsymbol{\Phi}}^{(0)})$$

то есть

$$(0 \cdot 0 + 1 \cdot 1 + 3 \cdot 3 + 6 \cdot 6) \cdot b_0 = (0.1 \cdot 0 + 0.8 \cdot 1 + 3.05 \cdot 3 + 5.95 \cdot 6)$$

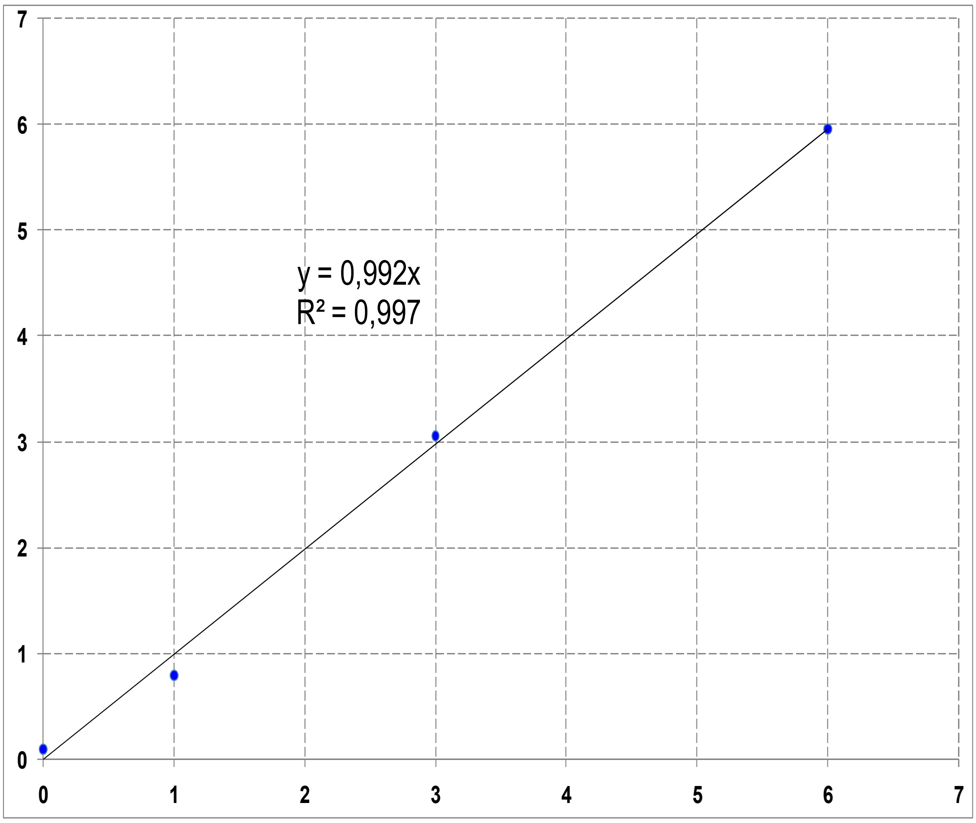
$$46 \cdot b_0 = 45.65$$

$$b_0 = \frac{45.65}{46} = 0.9923913$$

Следовательно,

$$\hat{Y} = 0.9923913 \cdot X$$

МНК-прямая (без константы) показана на Рисунке 5.



Р и с у н о к 5

П р и м е р 4

В таблице приведены результаты замеров:

$X_i^{(1)}$	0	1	3	6
$X_i^{(2)}$	2	5	7	8
Y_i	2.1	6.03	9.98	14.02

Нужно построить линейную зависимость Y от $X=(X^{(1)},X^{(2)})$ (без константы) методом наименьших квадратов. Указать, какой функционал будет минимизирован.

Р е ш е н и е

1) Количество наблюдений $n=4$. Обобщенный полином запишем в виде $\hat{Y}=b_0X^{(1)}+b_1X^{(2)}$. Считаем, что $\varphi_0(X)=X^{(1)}$, $\varphi_1(X)=X^{(2)}$ и $K=1$.

Параметры полинома должны быть решением задачи оптимизации

$$S(b_0)=\sum_{i=1}^{4}[\,Y_i-(b_0X_i^{(1)}+b_1X_i^{(2)})]^{\,2}\to\min$$

В данной задаче

$$S(b_0, b_1) = [2.1 - (b_0 \cdot 0 + b_1 \cdot 2)]^2 + [6.03 - (b_0 \cdot 1 + b_1 \cdot 5)]^2 +$$
$$+ [9.98 - (b_0 \cdot 3 + b_1 \cdot 7)]^2 + [14.02 - (b_0 \cdot 6 + b_1 \cdot 8)]^2$$

Решение задачи можно найти, записывая и решая нормальную систему уравнений

$$\frac{\partial S}{\partial b_0} = 0, \quad \frac{\partial S}{\partial b_1} = 0.$$

Используя (14.30), нормальную систему уравнений можно выписать быстрее.

Вектор истинных значений отклика $\overline{\boldsymbol{Y}}$ и матрица регрессоров $\boldsymbol{\Phi}$ размерности 4×2 равны

$$\overline{\boldsymbol{Y}} = \begin{bmatrix} 2.1 \\ 6.03 \\ 9.98 \\ 14.02 \end{bmatrix} \qquad \boldsymbol{\Phi} = \begin{bmatrix} \varphi_0(X_1) & \varphi_1(X_1) \\ \varphi_0(X_2) & \varphi_1(X_2) \\ \varphi_0(X_3) & \varphi_1(X_3) \\ \varphi_0(X_4) & \varphi_1(X_4) \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ 1 & 5 \\ 3 & 7 \\ 6 & 8 \end{bmatrix}$$

Здесь $X_1 = (0, 2), X_2 = (1, 5), X_3 = (3, 7), X_4 = (6, 8)$.

Столбцы матрицы регрессоров соответствуют векторам

$$\overline{\boldsymbol{\Phi}}^{(0)} = \begin{bmatrix} 0 \\ 1 \\ 3 \\ 6 \end{bmatrix} \qquad \overline{\boldsymbol{\Phi}}^{(1)} = \begin{bmatrix} 2 \\ 5 \\ 7 \\ 8 \end{bmatrix}$$

Нормальная система уравнений для отыскания $b_j,\ j = 0, 1$ имеет вид:

$$\begin{bmatrix} 0+1+9+36 & 2\cdot 0+5\cdot 1+7\cdot 3+8\cdot 6 \\ 2\cdot 0+5\cdot 1+7\cdot 3+8\cdot 6 & 4+25+49+64 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} =$$
$$= \begin{bmatrix} 2.1\cdot 0+6.03\cdot 1+9.98\cdot 3+14.02\cdot 6 \\ 2.1\cdot 2+6.03\cdot 5+9.98\cdot 7+14.02\cdot 8 \end{bmatrix}$$

то есть

$$\begin{bmatrix} 46 & 74 \\ 74 & 142 \end{bmatrix} \cdot \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \ldots$$

СЛАУ нужно решить самостоятельно.

Критерии качества МНК-приближения

Основными критериями для проверки качества МНК-полинома являются:

1) **величина функционала** S, то есть сумма квадратов остатков

$$S = \sum_{i=1}^{n} [\hat{\varepsilon}_i]^2 = \sum_{i=1}^{n} [Y_i - \hat{Y}_i]^2 \qquad (14.33)$$

(чем она меньше, тем лучше).

2) **коэффициент детерминации** R^2, его определением служит формула

$$R^2 = \frac{\sum_{i=1}^{n} [\hat{Y}_i - Y_{сред.}]^2}{\sum_{i=1}^{n} [Y_i - Y_{сред.}]^2} \qquad (14.34)$$

где

$$Y_{сред.} = \frac{\sum_{i=1}^{n} Y_i}{n} = \frac{\sum_{i=1}^{n} \hat{Y}_i}{n}$$

Областью значений коэффициента R^2 является отрезок $[0; 1]$. Чем ближе R^2 к единице, тем лучше.

3) величина s, ее называют **«стандартная ошибка оценки»**, определяют формулой

$$s = \sqrt{\frac{\sum_{i=1}^{n} [\hat{\varepsilon}_i]^2}{n - (K+1)}} \qquad (14.35)$$

Здесь через $K+1$ обозначено число параметров, которые должны быть найдены при построении МНК-полинома. Чем меньше s, тем лучше.

4) визуальный и статистический анализ остатков $\hat{\varepsilon}_i, \quad i = 1,...n$. Например, графики зависимости остатков от фактора X или отклика Y, а также нормальный вероятностный график остатков.

Если графики остатков хаотичны, значит, функциональная зависимость отклика Y от фактора X описывается МНК-полиномом достаточно полно (не учтено и не может быть учтено только случайное поведение Y).

Если графики остатков показывают какую-либо закономерность, желательно пересмотреть аппроксимирующее уравнение (полином, обобщенный полином) и включить в него недостающие компоненты.

Анализ качества построенного МНК-полинома (обощенного полинома) начинают с анализа остатков.