KARPOV.COURSES >>>
КОНСПЕКТ

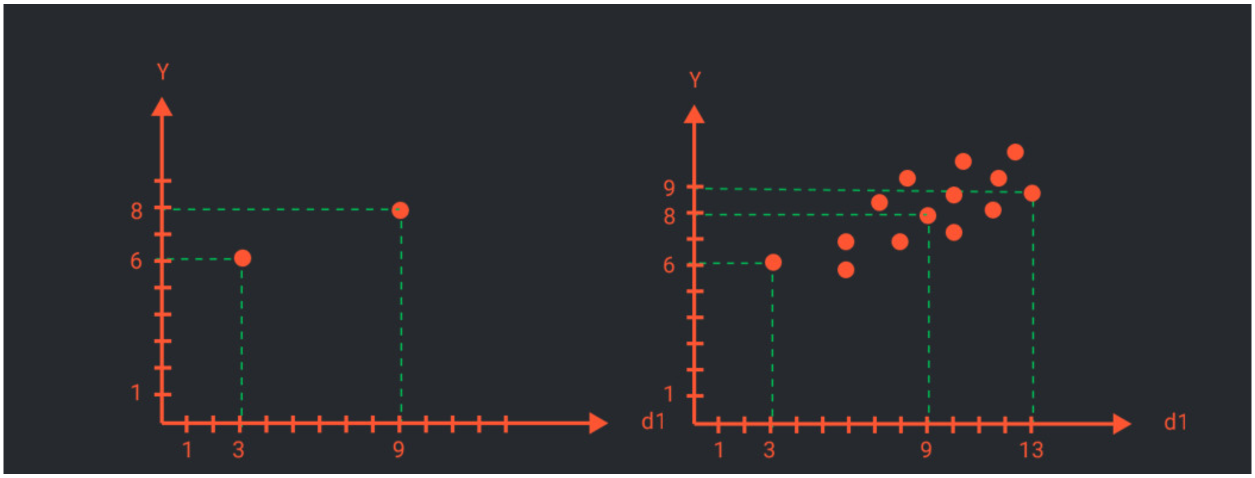
Конспект > 3 урок > Линейная регрессия

>Линейные модели

Линейная модель(linear model) - модель, отображающая состояние или функционирование системы таким образом, что все взаимозависимости в ней являются линейными и могут быть выражены уравнением:

$$a(x) = \beta_0 + \beta_1 \cdot d_1 + \beta_2 \cdot d_2 + ... + \beta_n \cdot d_n$$

, где d - признаки нашего объекта, β - некие числовые коэффициенты, которые также принято называть весами и параметрами

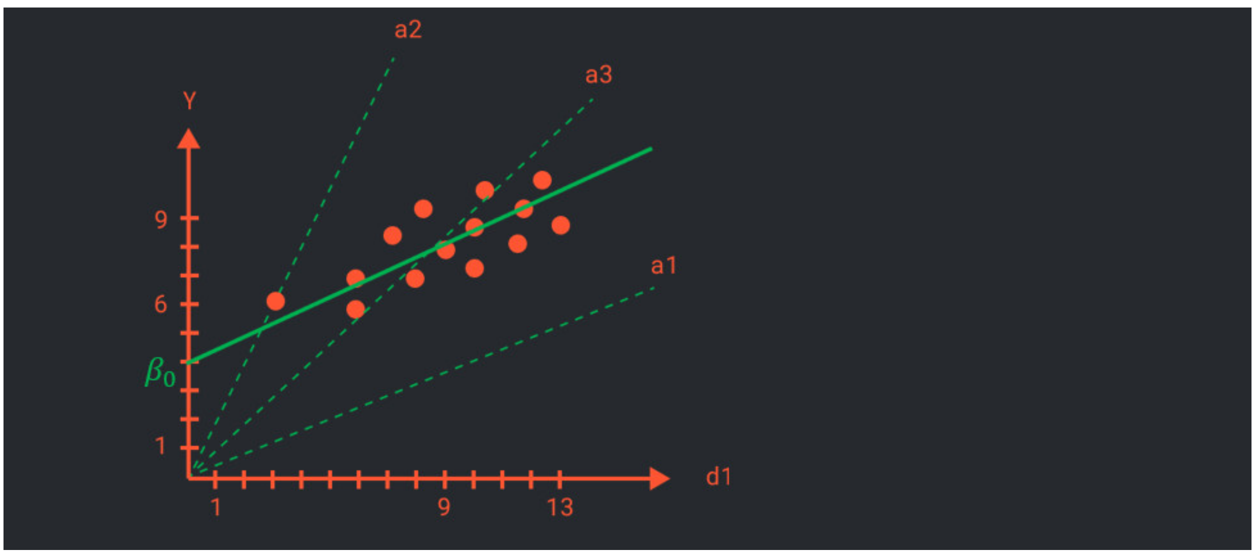


$a_1(x) = 0,5 \cdot d_1$

$a_2(x) = 2 \cdot d_1$

$a_3(x) = 1 \cdot d_1$

Как можно убедится на графике зависимость между $a(x)$ и d_1 можно выразить через линию, а от коэффициента β будет зависеть степерь ее наклона к оси OX.



Мы получили первое семейство моделей :)
Ближе всех описывает поведение наших точек график 3.

Кроме коэффициента
β, который влияет лишь на степень наклона, можно двигать линию вдоль оси OY с помощью так назваемого свободного коэффициента β_0.

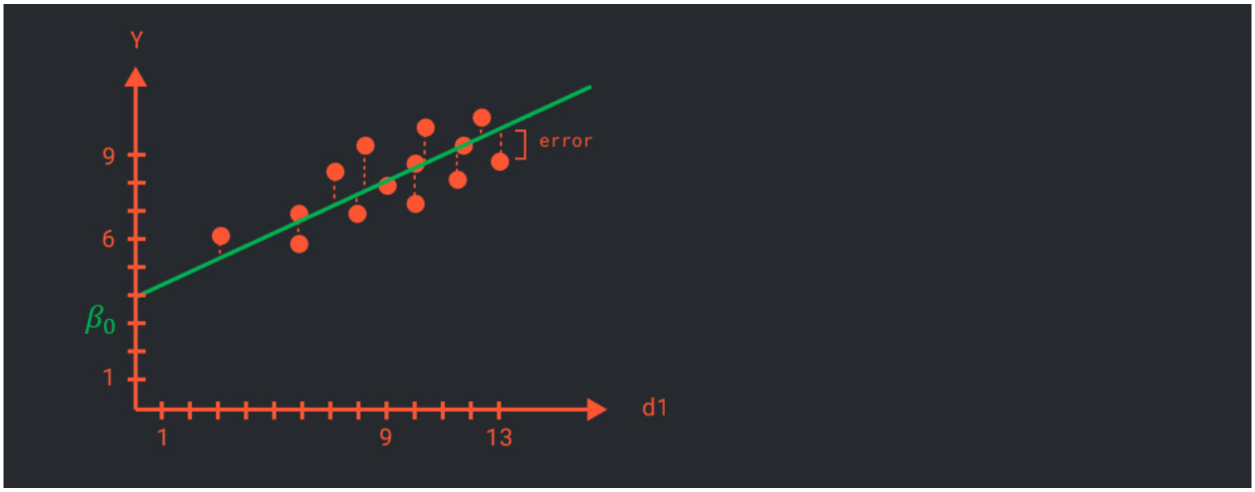
Предположим у нас есть набор объектов с единственным известным признаком d_1:

Мы можем отразить данную зависимость на графике:

Задача МО - найти такие значения коэффициентов β, при которых зависимость между объектами X и таргетными значениями Y будет отражена наиболее правдоподобно.

При правильно выбранном Losse, это равноситьно поиску минимума функционала качества:

$\beta^* = argmin(Q(a(x), X))$, то есть такие параметры, при которых функционал качества будет минимальным. Графически это можно выразить как минимизацию суммы всех отклонений точек от проведенной прямой:



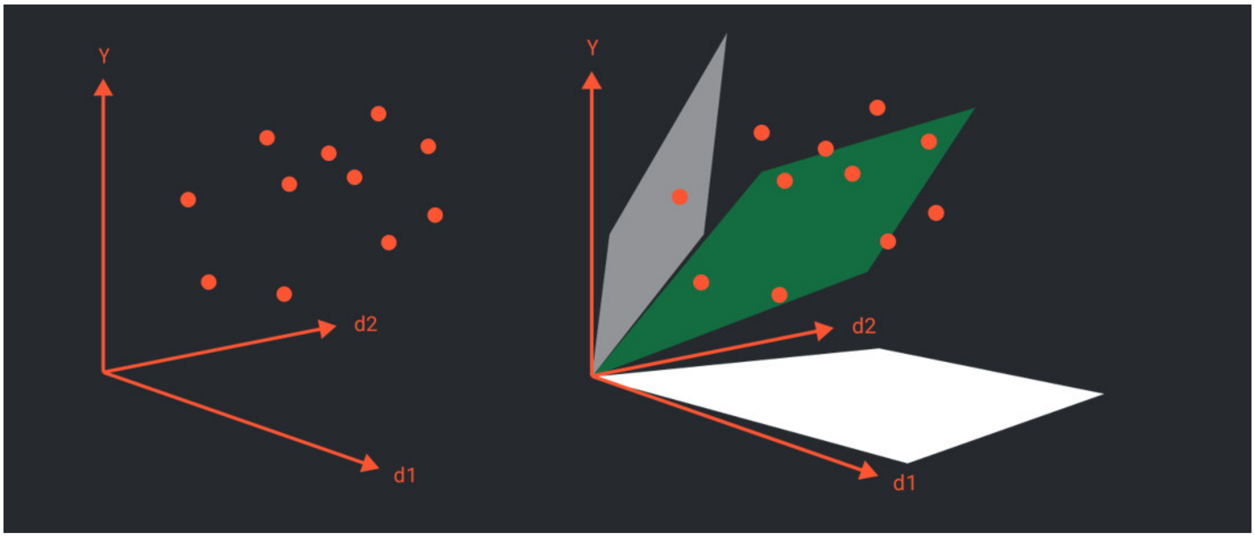
>Многомерная линейная регрессия

Теперь представим, что у нашего объекта не один, а несколько признаков:

$$x_1 = (d_1^1, d_2^1) = (3, 2) \qquad y_1 = 6$$

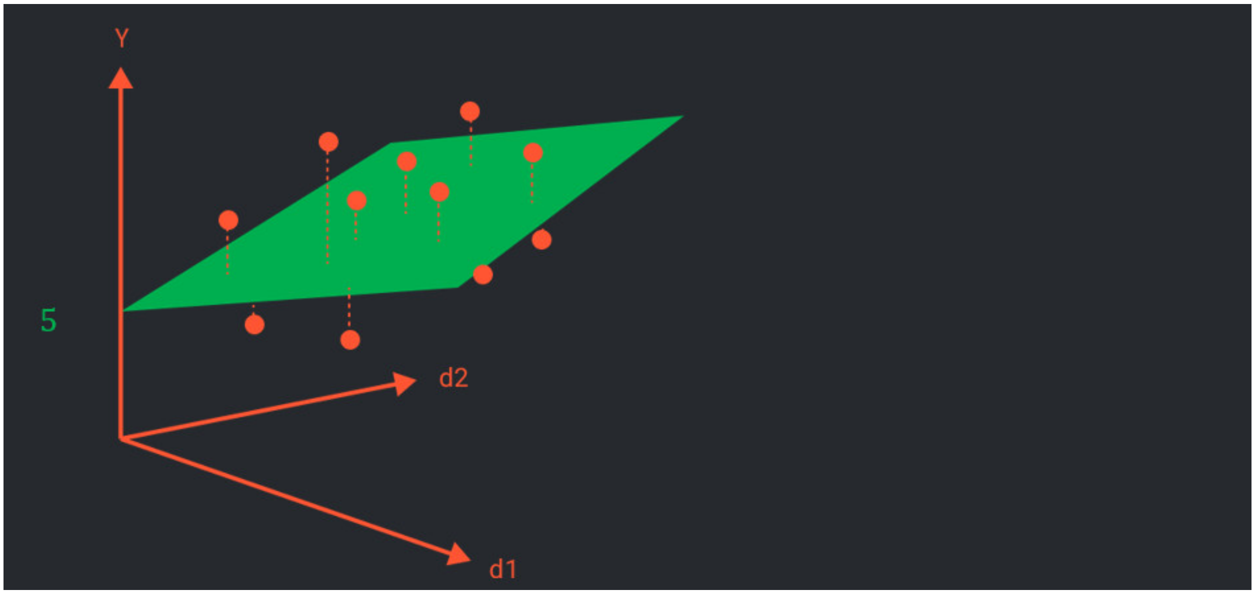
$$x_2 = (d_1^2, d_2^2) = (9, 5) \qquad y_2 = 8$$

Тогда зависимость будет отражена уже в трехмерном пространстве, и для того, чтобы ее описать нам понадобится провести уже не прямую, а плоскость, которая будет проходить максимально близко ко всем точкам.



$$a(x) = \beta_1 d_1 + \beta_2 d_2 + \beta_0$$

И снова задача подобрать те параметры, при подстановке которых в функцию, функционал качества будет минимальным и соответственно минимальное среднее всех расстояний до плоскости (длин проведенных к плоскости перепендикуляров).



>Линейная регрессия OLS

Для поиска лучших коэффициентов β используется функционал среднеквадратичной ошибки, уже упомянутый выше:

$$Q(a(x), X) = \frac{1}{m}\sum_{i=1}^{m} L(a(x_i), y_i)$$

В качестве примера рассмотрим задачу по предсказанию цены акции компании в USD через год (y_i) по текущей цене (d_1) и кредиторской задолженности в миллионах (d_2). Допустим в выборку попало три компании (x_1, x_2, x_3).

$$\begin{array}{c c} & \begin{array}{cc} d_1 & d_2 \end{array} \\ \begin{array}{c} x_1 \\ x_2 \\ x_3 \end{array} & \begin{pmatrix} 23 & 0{,}5 \\ 35 & 1 \\ 18 & 0 \end{pmatrix} \end{array} \qquad \begin{pmatrix} y_1 & 55 \\ y_2 & 100 \\ y_3 & 45 \end{pmatrix}$$

1. Построим линейную модель для данной зависимости, подставив имеющиеся признаки в уравнение общего вида

$$a(x) = \beta_1 d_1 + \beta_2 d_2 + \beta_0\ .$$

Для первого объекта: $a(x_1) = \beta_1 \cdot 23 + \beta_2 \cdot 0.5 + \beta_0$

Для второго объекта: $a(x_2) = \beta_1 \cdot 35 + \beta_2 \cdot 1 + \beta_0$

Для третьего объекта: $a(x_3) = \beta_1 \cdot 18 + \beta_2 \cdot 0 + \beta_0$

2. Теперь подставим значения в функционал качества:

$MSE = \frac{1}{N}\sum_{i=1}^{N}((a(x_i) - y_i))^2 = \dfrac{1}{3}((a(x_1) - y_1))^2 + ((a(x_2) - y_2))^2 + ((a(x_3) - y_3))^2$

3. Подставим значения $a(x)$, полученные на первом шаге в функцию MSE:

$MSE = \frac{1}{3}((\beta_1 \cdot 23 + \beta_2 \cdot 0.5 + \beta_0 - y_1))^2 + ((\beta_1 \cdot 35 + \beta_2 \cdot 1 + \beta_0 - y_2))^2 + ((\beta_1 \cdot 18 + \beta_2 \cdot 0 + \beta_0 - y_3))^2$

4. Итак мы получили итоговую функцию, которую нужно минимизировать. Для этого нужно обратиться к понятию производной функции.

>Ликбез №2: Производные и экстремумы функции

Производная показывает, как изменяется функция с изменением аргумента:

$$y' = \frac{\Delta y}{\Delta x}$$

где Δ - (delta) величина изменения, при этом $\Delta \approx 0$

можно раскрыть это уравнение и представить как $y' = \frac{\Delta y(x+\Delta(x)) - y(x)}{\Delta x}$

, где $\Delta(x) \to 0$

Если производная функции положительна, то функция возрастает и наоборот. Точки, в которых производная равна нулю называют критическими точками или точками экстремума.

Дифференцирование - взятие производной функции

Экстрем́ум - максимальное или минимальное значение функции на заданном множестве.

Максимум функции - точка, где возрастание функции сменяется убыванием.

Минимум - точка, где убывание функции сменяется возрастанием.

Максимум и минимум могут быть локальные и глобальные.

В точках экстремума производная функции равна **нулю**.

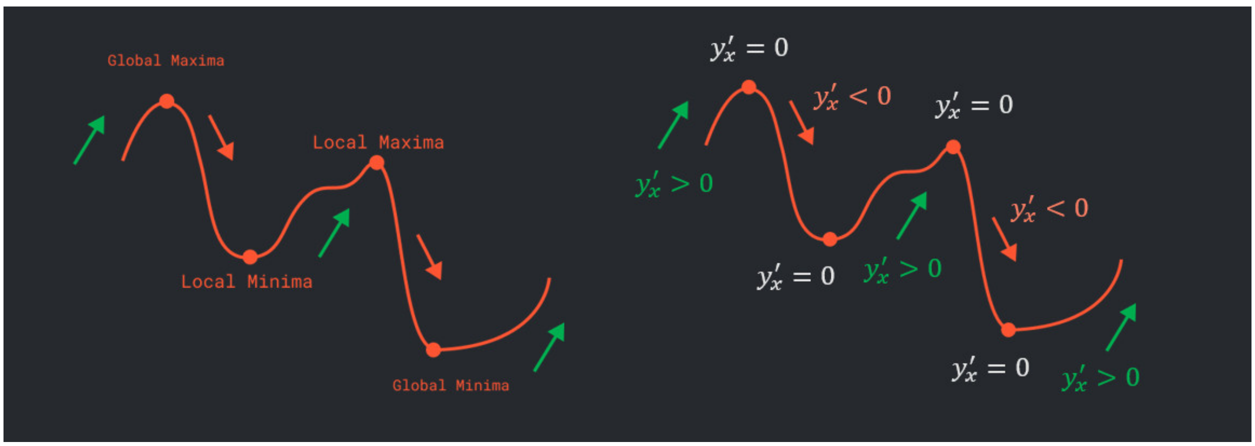


Таблица производных

$(x^n)' = n \cdot x^{n-1}$	$(\sin x)' = \cos x$	$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}$
$(a^x)' = a^x \cdot \ln(a)$	$(\cos x)' = -\sin x$	$(\arccos x)' = \frac{-1}{\sqrt{1-x^2}}$
$(\log_a x)' = \frac{1}{x \cdot \ln a}$	$(\operatorname{tg} x)' = \frac{1}{\cos^2 x}$	$(\operatorname{arctg} x)' = \frac{1}{1+x^2}$
$(\sin x)' = \cos x$	$(\operatorname{ctg} x)' = -\frac{1}{\sin^2 x}$	$(\operatorname{arcctg} x)' = \frac{-1}{1+x^2}$

Правила вычисления производных

Геометрический смысл производной видео

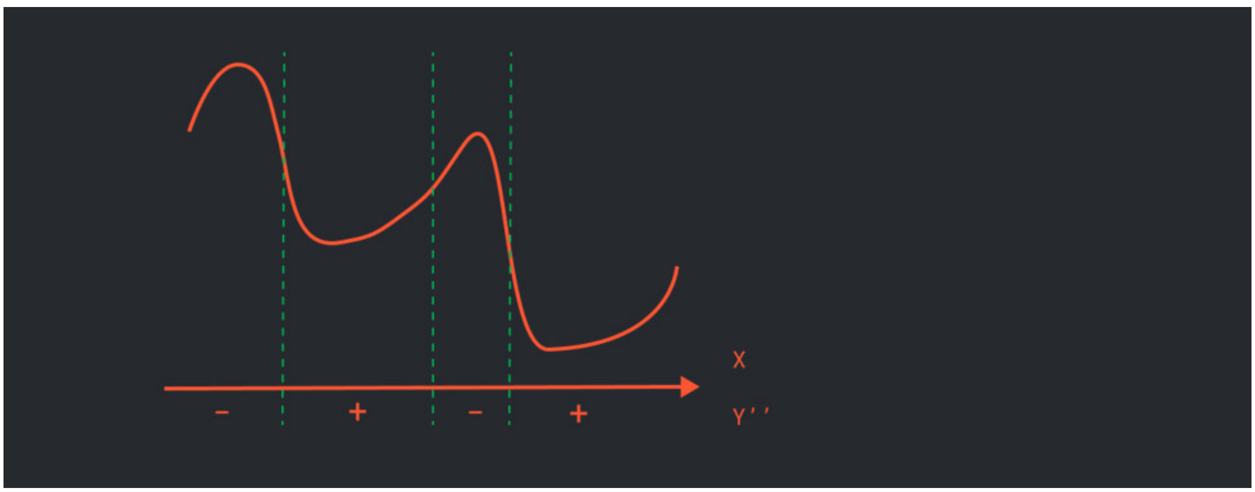
>Условие второго порядка - выпуклость и вогнутость функции

Для того, чтобы определить, минимумом или максимумом является найденная точка экстремума, используется вторая производная - производная от производной:

$$y'' = (y')'$$

С помощью нее можно оценить выпуклость и вогнутость функции.

Если вторая производная в точке положительная, то мы нашли минимум и наоборот.



В качестве примера исследуем на экстремумы следующую функцию :

$$y = x^3 - 6x^2 - 15x + 10$$

1. Запишем уравнение производной функции:

$$y' = (x^3 - 6x^2 - 15x + 10)' = (x^3)' - 6(x^2)'(10)' = 3x^2 - 12x - 15 = 0$$

Решив уравнение получим две критические точки:

$$x_1 = -1 \qquad x_2 = 5$$

2. Исследуем функцию на возрастание и убывание на промежутках между найденными точками.

Для этого найдем знак производной каждом из участков, подставив в уравнения произвольные значения x , принадлежащие каждому из участков:

$y'(-2) = 3 \cdot (-2)^2 - 12x - 15 = 21 > 0$

$y'(0) = 0 - 0 - 15 = -15 < 0$

$y'(6) = 3 \cdot (6)^2 - 12 \cdot 6 - 15 = 21 > 0$

3. Другой способ найти минимум и максимум, посчитать вторые производные:

$y''(-1) = (3x^2 - 12x - 15)' = 6x - 12 = -18 < 0 \Rightarrow$ Локальный максимум

$y''(5) = (3x^2 - 12x - 15)' = 6x - 12 = 18 > 0 \Rightarrow$ Локальный минимум

>Производная функции нескольких переменных

Для поиска минимума функции с несколькими переменными нам понадобится обратиться к частной производной - производной функции нескольких переменных. Сам алгоритм нахождения минимума остается тем же.

Рассмотрим функцию $z(x, y) = x2 + 3y2$

Шаг1: Находим производные первого порядка по каждому аргументу:

$$\begin{cases} z_x\prime = 2 \cdot x = 0 \\ z_y\prime = 6 \cdot y = 0 \end{cases}$$

Шаг2: Находим частные производные второго порядка:

$z_{xx}\prime\prime = (zx\prime)x\prime = 2$

$$z_{yx}'' = (zy')x' = 0$$

$$z_{xy}'' = (zx')y' = 0$$

$$z_{yy}'' = (zy')y' = 6$$

Шаг3: Теперь проверяем функцию на выпуклость /вогнутость. Для этого найденные на предыдущем шаге значения вторых производных подставляем в матрицу Гессе (матрица вторых частных производных функции нескольких переменных):

$$H_z(x^*, y^*) = \begin{pmatrix} 2 & 0 \\ 0 & 6 \end{pmatrix}$$

Далее используют критерий Сильвестра, который не разбирается в данном курсе, так как среднеквадратичная функция ошибки представляет собой параболу и всегда будет выпукла вниз, а значит найденная точка экстремума и будет искомой точкой минимума.

Разобраться с производными более детально поможет эта статья.

>Линейная регрессия OLS: пример

По сути функционал качества является функцией многих переменных.

Теперь мы можем его минимизировать, используя знания о производной и экстремах функции:

Составим систему из частных производных от функции среднеквадратичной ошибки по каждому из трех параметров:

$$Q = \frac{1}{3} \cdot [\,(\beta_1 \cdot 23 + \beta_2 \cdot 0.5 + \beta_0 - 55)^2 + (\beta_1 \cdot 35 + \beta_2 \cdot 1 + \beta_0 - 100)^2 + (\beta_1 \cdot 18 + \beta_0 - 45)^2] \to min$$

$$\begin{cases} Q'_{\beta_1} = \frac{1}{3} \cdot [23 \cdot 2 \cdot (\beta_1 \cdot 23 + \beta_2 \cdot 0.5 + \beta_0 - 55) + 35 \cdot 2 \cdot (\beta_1 \cdot 35 + \beta_2 \cdot 1 + \beta_0 - 100) + 18 \cdot 2 \cdot (\beta_1 \cdot 18 + \beta_0 - 45)] = 0 \\ Q'_{\beta_2} = \frac{1}{3} \cdot [0.5 \cdot 2 \cdot (\beta_1 \cdot 23 + \beta_2 \cdot 0.5 + \beta_0 - 55) + 1 \cdot 2 \cdot (\beta_1 \cdot 35 + \beta_2 \cdot 1 + \beta_0 - 100)] = 0 \\ Q'_{\beta_0} = \frac{1}{3} \cdot [1 \cdot 2 \cdot (\beta_1 \cdot 23 + \beta_2 \cdot 0.5 + \beta_0 - 55) + 1 \cdot 2 \cdot (\beta_1 \cdot 35 + \beta_2 \cdot 1 + \beta_0 - 100) + 1 \cdot 2 \cdot (\beta_1 \cdot 18 + \beta_0 - 45)] = 0 \end{cases}$$

Решив эту систему, получим критическую точку с параметрами:

$$\beta_1^*, \beta_2^*, \beta_0^* = (5, -30, -45)$$

Далее проверкой на выпуклость убеждаемся, что нашли минимум (спойлер: OLS функция **всегда выпукла вниз**) и эти значения подставляем в итоговую формулу:

$$a(x) = 5 \cdot d_1 - 30 \cdot d_2 - 45$$

Таким образом мы нашли формулу лучшей описывающей линейной модели для поставленнной задачи.

Однако такое решение на практике является довольно громоздким. Есть способ для более удобной работы с задачами линейной регрессии - применение матриц.

>Ликбез №3: Матрицы

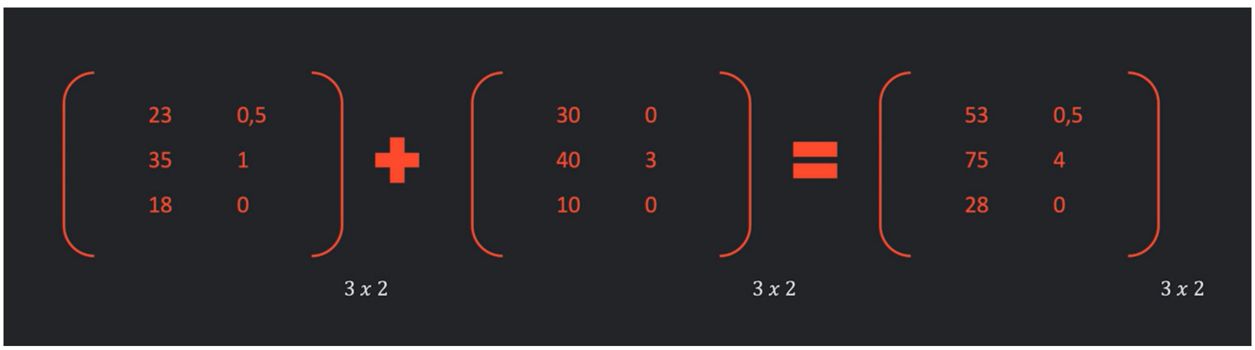
Матрица – форма хранения чисел в математике, более точно упорядоченный массив элементов в виде прямоугольной таблицы (`m` x `k`).

Вектор – матрица с только одним столбцом или одной строкой (`n` x `1`, `1` x `n`).

Матрицы можно складывать, вычитать, умножать.

Матричное сложение и вычитание

Осуществляется поэлементно, сложение определено (возможно) только для матриц одинаковой размерности.



Матричное умножение

Матричное умножение `A·B` возможно (определено), если `A` имеет размер (`n` x `k`), а `B` - (`k` x `m`), то есть когда количество столбцов первой матрицы совпадает с количеством строк второй.

Чтобы получить c_{ij}, нужно скалярно перемножить `i` строку `A` и `j` столбец `B`:

$$\underset{3\,x\,2}{\overset{A}{\begin{pmatrix} 23 & 0{,}5 \\ 35 & 1 \\ 18 & 0 \end{pmatrix}}} \underset{2\,x\,1}{\overset{B}{\begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix}}} = \underset{3\,x\,2}{\overset{C}{\begin{pmatrix} 23 \cdot \beta_1 + 0{,}5 \cdot \beta_2 \\ 35 \cdot \beta_1 + 1 \cdot \beta_2 \\ 18 \cdot \beta_1 + 0 \cdot \beta_2 \end{pmatrix}}}$$

Важно помнить, что в отличии от привычных правил арифметики, для матриц от перестановки сомножителей результат меняется!

Транспонирование матрицы

Транспонировать матрицу — значит превратить ее строки в столбцы и наоборот. При этом она как бы ложится на бок:

$$\underset{3\,x\,2}{\overset{A}{\begin{pmatrix} 23 & 0{,}5 \\ 35 & 1 \\ 18 & 0 \end{pmatrix}}} \xrightarrow{T} \underset{2\,x\,3}{\overset{A^{\mathrm{T}}}{\begin{pmatrix} 23 & 35 & 18 \\ 0{,}5 & 1 & 0 \end{pmatrix}}}$$

Единичная матрица

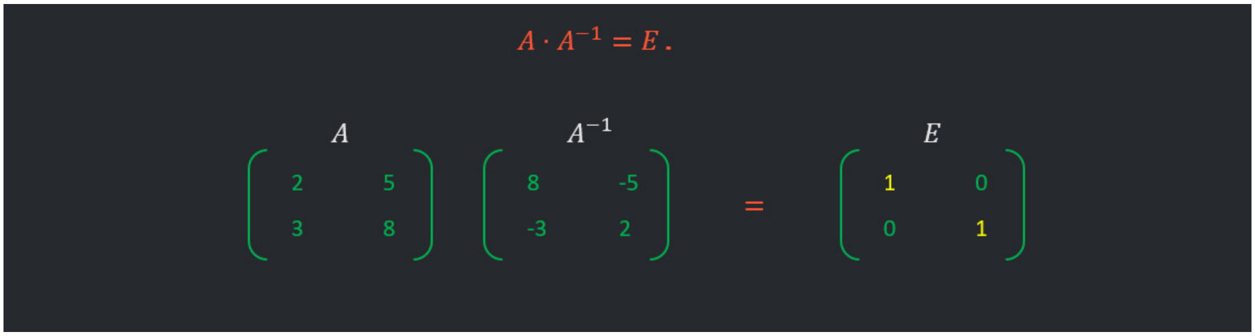
Единичная матрица — квадратная матрица, диагональные значения которой равны единице, остальные нулю. Обозначается символом `E` и является аналогом единицы в мире матриц.

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

Обратная матрица

Матрица, при умножении на которую исходной матрицы получается единичная матрица:

$$A^{-1} \cdot A = E$$



Не все матрицы являются обратимыми. Одним из условий обратимости матрицы является то, что она должна быть квадратной.

> Решение задачи регрессии OLS: матричная форма

Второй способ решить ту же задачу линейной регрессии — использовать матричную форму.

Представим выборку из объектов и признаков в виде матрицы X размера n х k,

где n — количество элементов в выборке, k — количество признаков, Y — вектор ответов (таргет), B — вектор коэффициентов β.

Тогда формулу среднеквадратичной ошибки, которую мы использовали ранее, можно записать в более компактном виде:

$$Q = \frac{1}{n}(X \cdot B - Y)^T(X \cdot B - Y) -- > min$$

Задача поиска оптимальных коэффициентов сводится к минимицации произведения двух матриц:

$$\frac{1}{3}\cdot\left(\begin{bmatrix}23&0,5&1\\35&1&1\\18&0&1\end{bmatrix}\begin{bmatrix}\beta_1\\\beta_2\\\beta_0\end{bmatrix}-\begin{bmatrix}55\\100\\45\end{bmatrix}\right)^T\cdot\left(\begin{bmatrix}23&0,5&1\\35&1&1\\18&0&1\end{bmatrix}\begin{bmatrix}\beta_1\\\beta_2\\\beta_0\end{bmatrix}-\begin{bmatrix}55\\100\\45\end{bmatrix}\right)=$$

$$\frac{1}{3}\cdot\left(\begin{bmatrix}23\cdot\beta_1+0,5\cdot\beta_2+\beta_0\\35\cdot\beta_1+\beta_2+\beta_0\\18\cdot\beta_1+\beta_0\end{bmatrix}-\begin{bmatrix}55\\100\\45\end{bmatrix}\right)^T\cdot\left(\begin{bmatrix}23\cdot\beta_1+0,5\cdot\beta_2+\beta_0\\35\cdot\beta_1+\beta_2+\beta_0\\18\cdot\beta_1+\beta_0\end{bmatrix}-\begin{bmatrix}55\\100\\45\end{bmatrix}\right)=$$

$$\frac{1}{3}\cdot\begin{pmatrix}23\cdot\beta_1+0,5\cdot\beta_2+\beta_0-55\\35\cdot\beta_1+\beta_2+\beta_0-100\\18\cdot\beta_1+\beta_0-45\end{pmatrix}^T\cdot\begin{pmatrix}23\cdot\beta_1+0,5\cdot\beta_2+\beta_0-55\\35\cdot\beta_1+\beta_2+\beta_0-100\\18\cdot\beta_1+\beta_0-45\end{pmatrix}=$$

$$\frac{1}{3}\cdot\left[(\beta_1\cdot 23+\beta_2\cdot 0.5+\beta_0-55)^2+(\beta_1\cdot 35+\beta_2\cdot 1+\beta_0-100)^2+(\beta_1\cdot 18+\beta_0-45)^2\right]$$

Далее для поиска минимума можно взять матричный дифференциал и приравнять к нулю. Этот шаг аналогичен вычислению частных производных и решению системы уравнений для поиска критических точек.

$$dQ(\beta^*) = 0$$

$$\beta^* = \begin{bmatrix}\beta_1^*\\\beta_2^*\\\beta_0^*\end{bmatrix} = (X^T\cdot X)^{-1}\cdot X^T\cdot Y = \begin{bmatrix}+5\\-30\\-45\end{bmatrix}$$

- В матричном способе решения используется дифференцирование матрицы, которое мы не рассматриваем в данном курсе.

Статья для желающих более подробно ознакомится с матричным методом.

Статья о линейной регрессии на Хабре.