# START ML

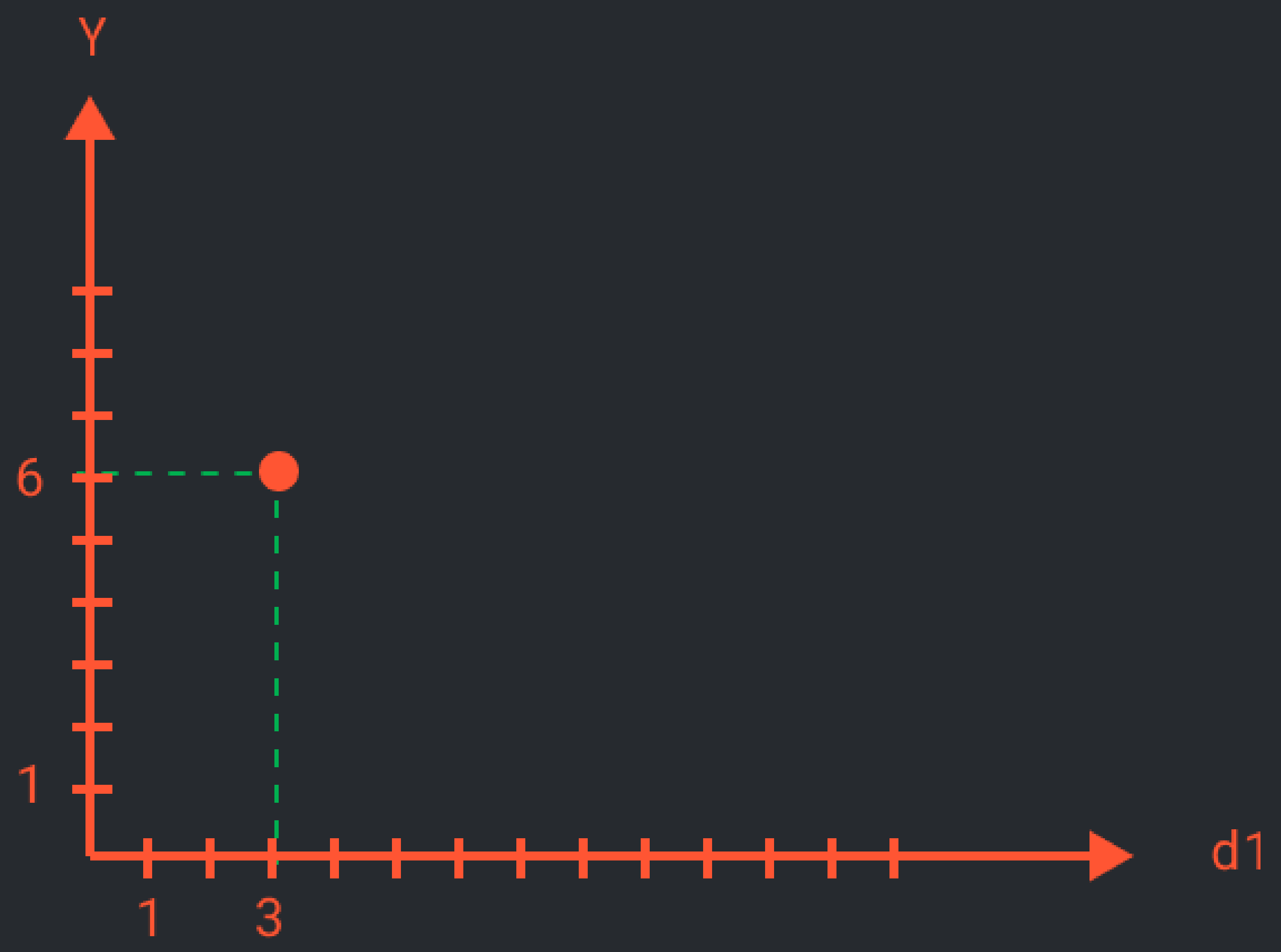
KARPOV.COURSES

# ЛИНЕЙНЫЕ МОДЕЛИ

Пусть есть много наборов $(x_i, y_i)$

Причем о наших объектах нам известен исключительно один признак $d_1$

$$x_1 = d_1^1 = 3 \qquad y_1 = 6$$

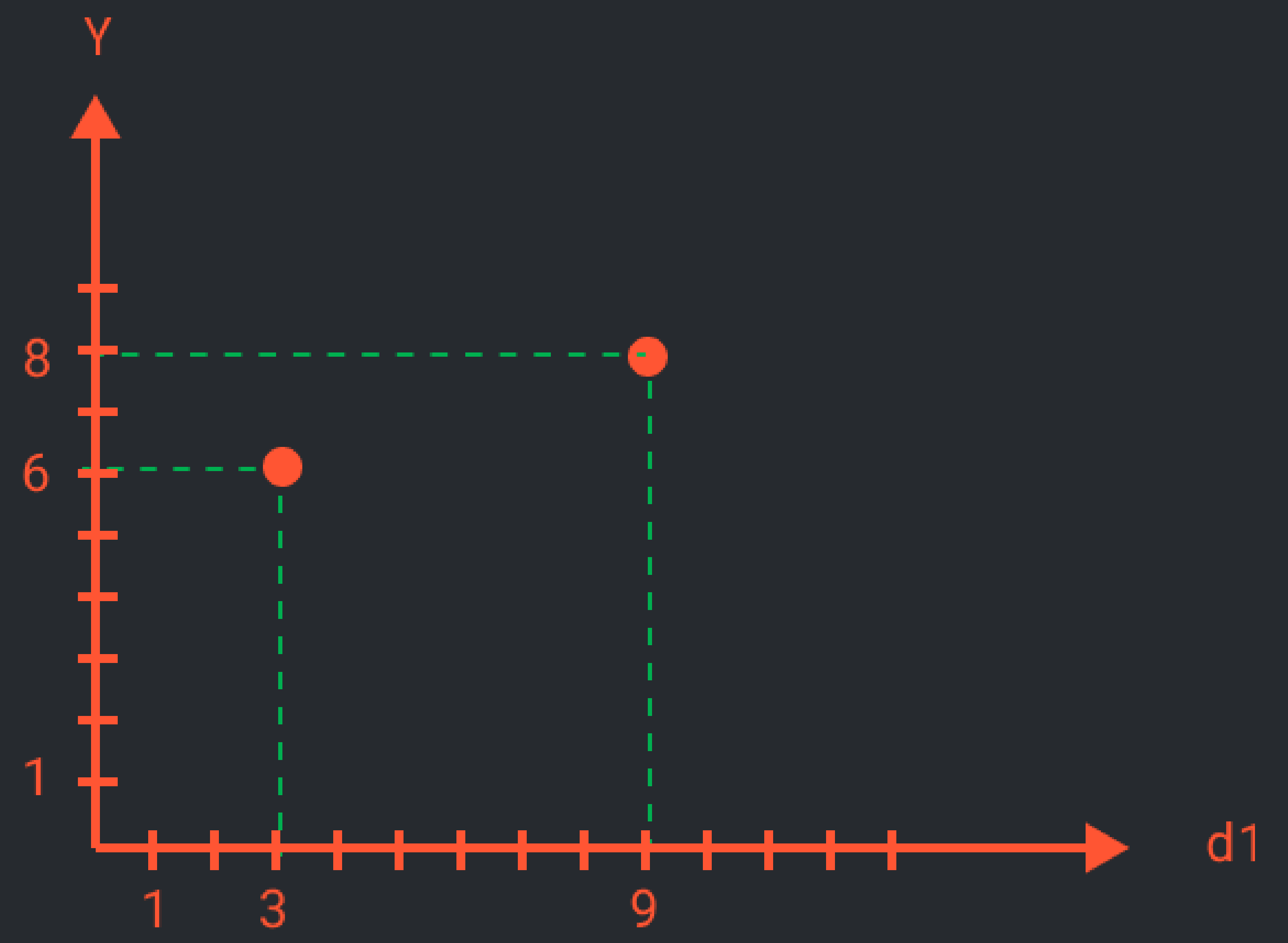


# ЛИНЕЙНЫЕ МОДЕЛИ

Пусть есть много наборов $(x_i, y_i)$

Причем о наших объектах нам известен исключительно один признак $d_1$

$$x_1 = d_1^1 = 3 \qquad y_1 = 6$$
$$x_2 = d_1^2 = 9 \qquad y_2 = 8$$

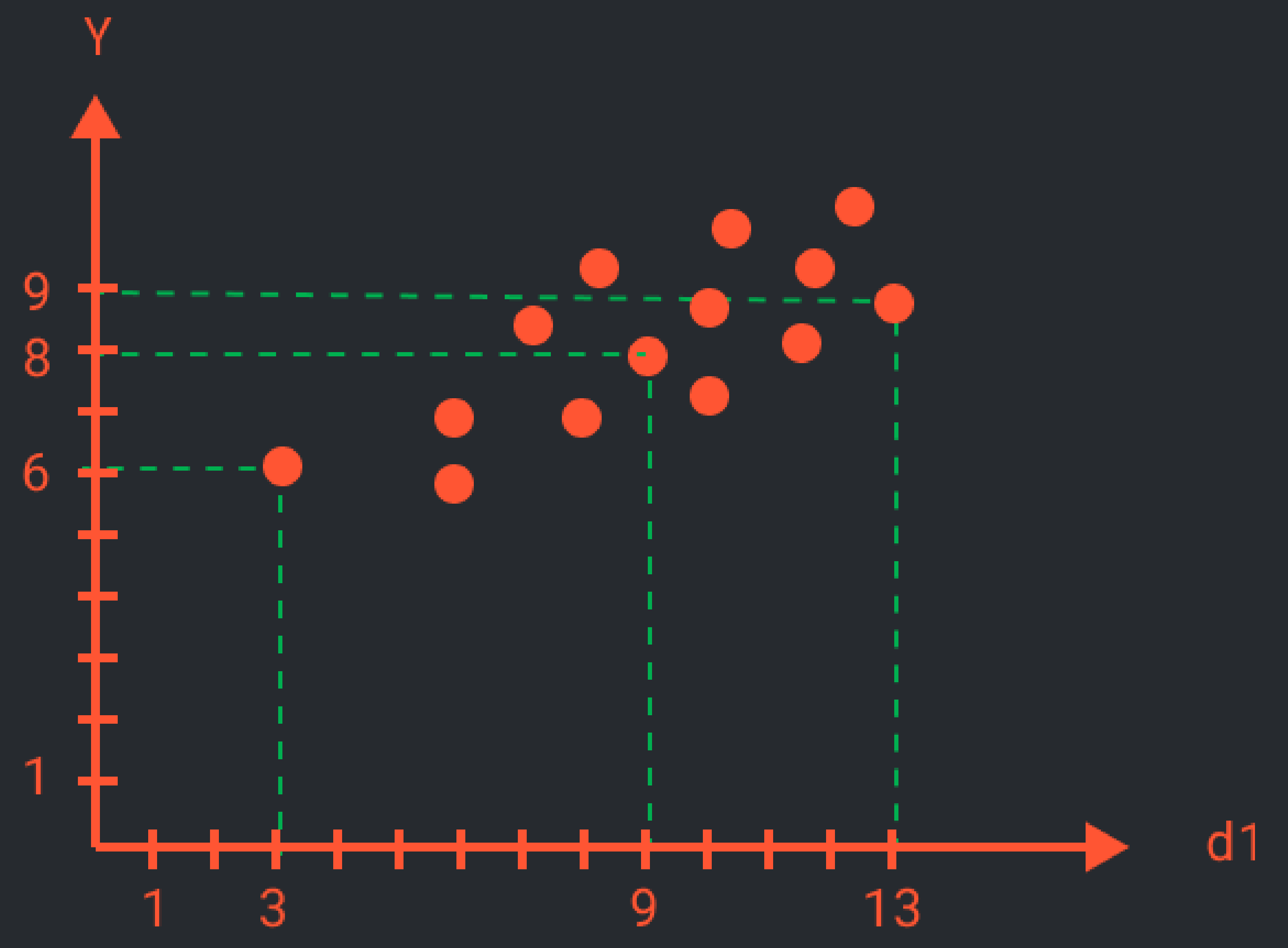


# ЛИНЕЙНЫЕ МОДЕЛИ

Пусть есть много наборов $(x_i, y_i)$

Причем о наших объектах нам известен исключительно один признак $d_1$

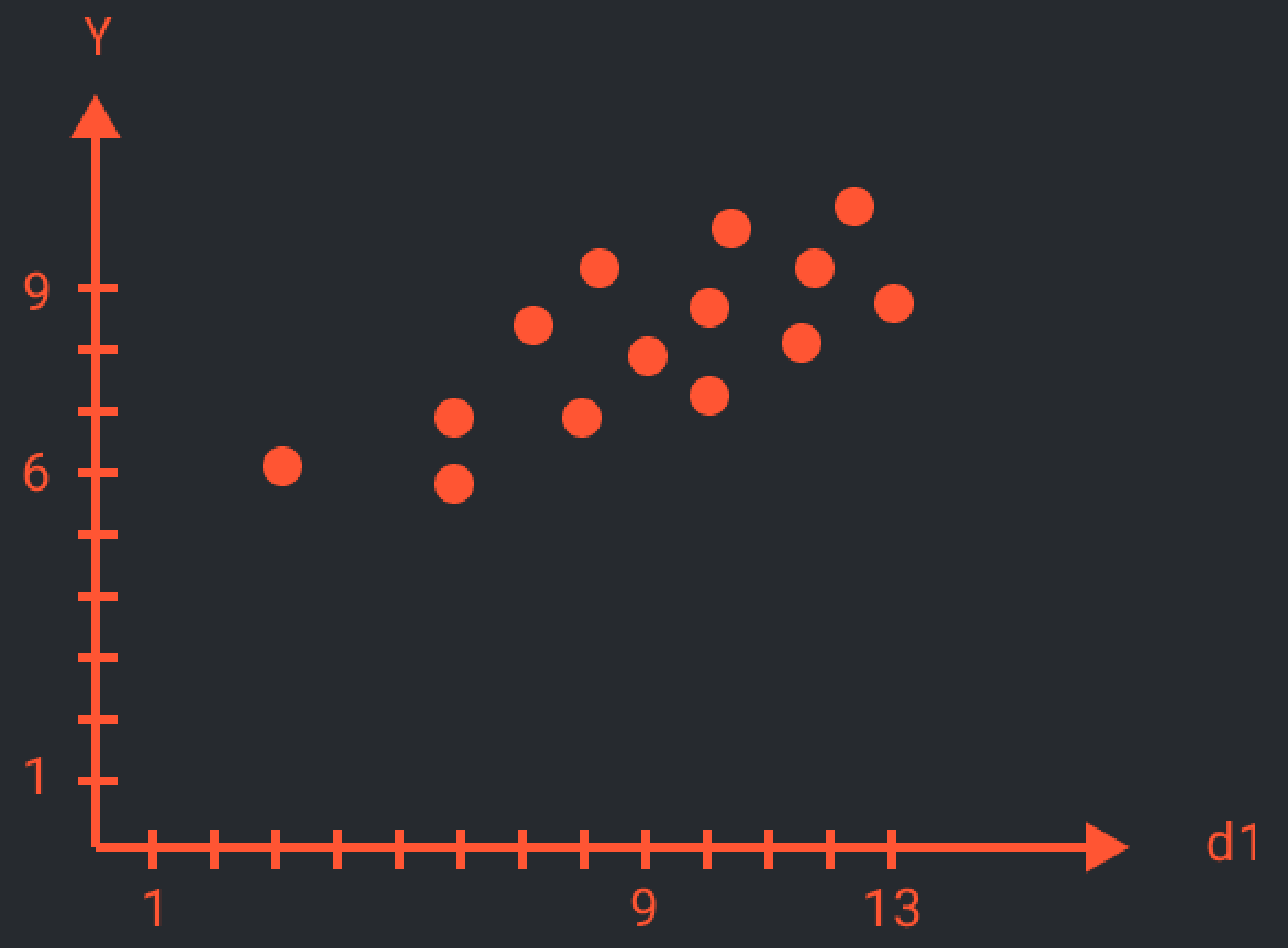
$$x_1 = d_1^1 = 3 \qquad y_1 = 6$$
$$x_2 = d_1^2 = 9 \qquad y_2 = 8$$
$$\dots$$
$$x_n = d_1^n = 13 \qquad y_n = 9$$



# ЛИНЕЙНЫЕ МОДЕЛИ

Построим нашу первую линейную модель!

$$a(x) = \beta_1 \cdot d_1$$

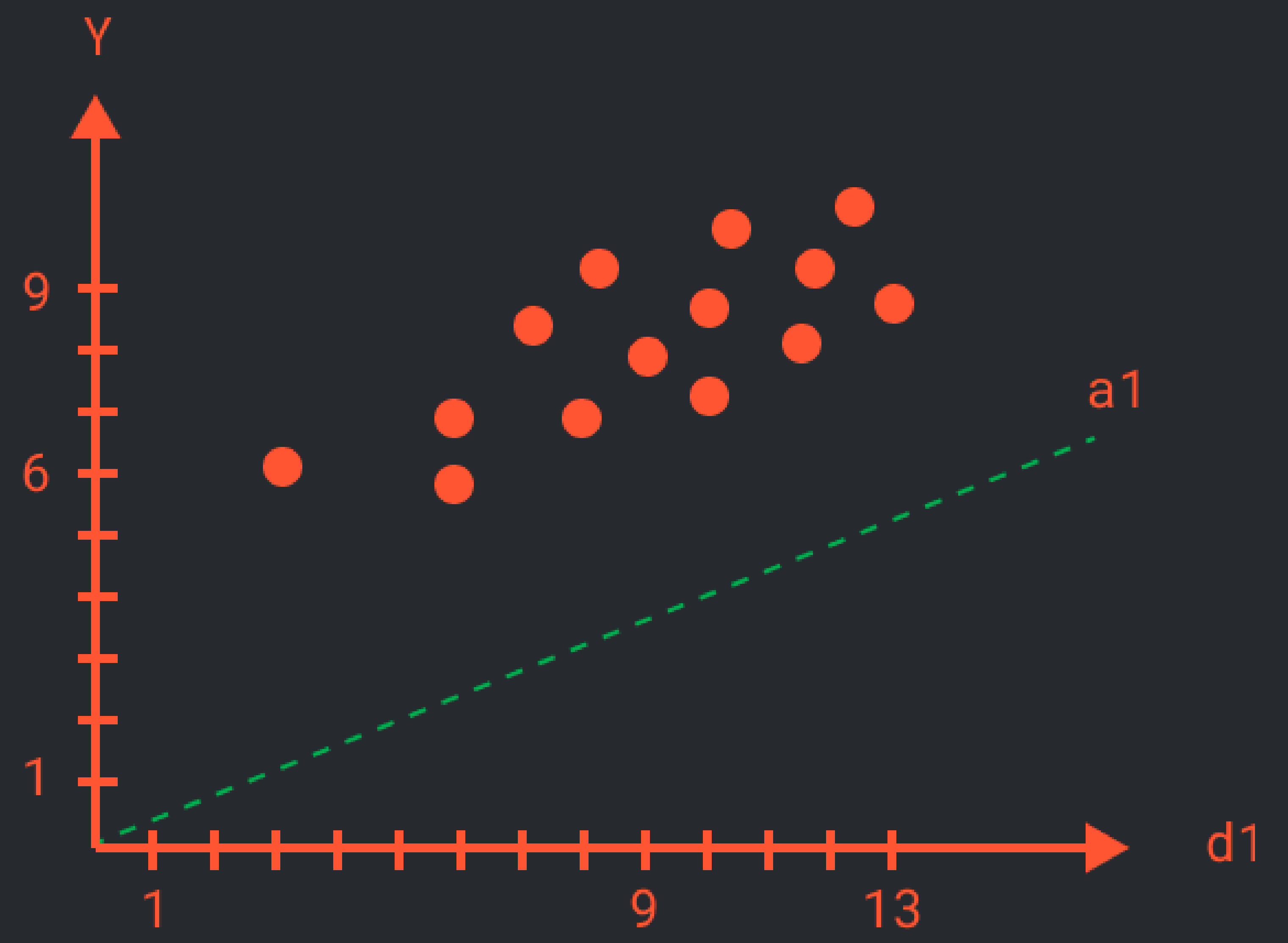


# ЛИНЕЙНЫЕ МОДЕЛИ

Построим нашу первую линейную модель!

$$a(x) = \beta_1 \cdot d_1$$

$$a_1(x) = 0.5 \cdot d_1$$

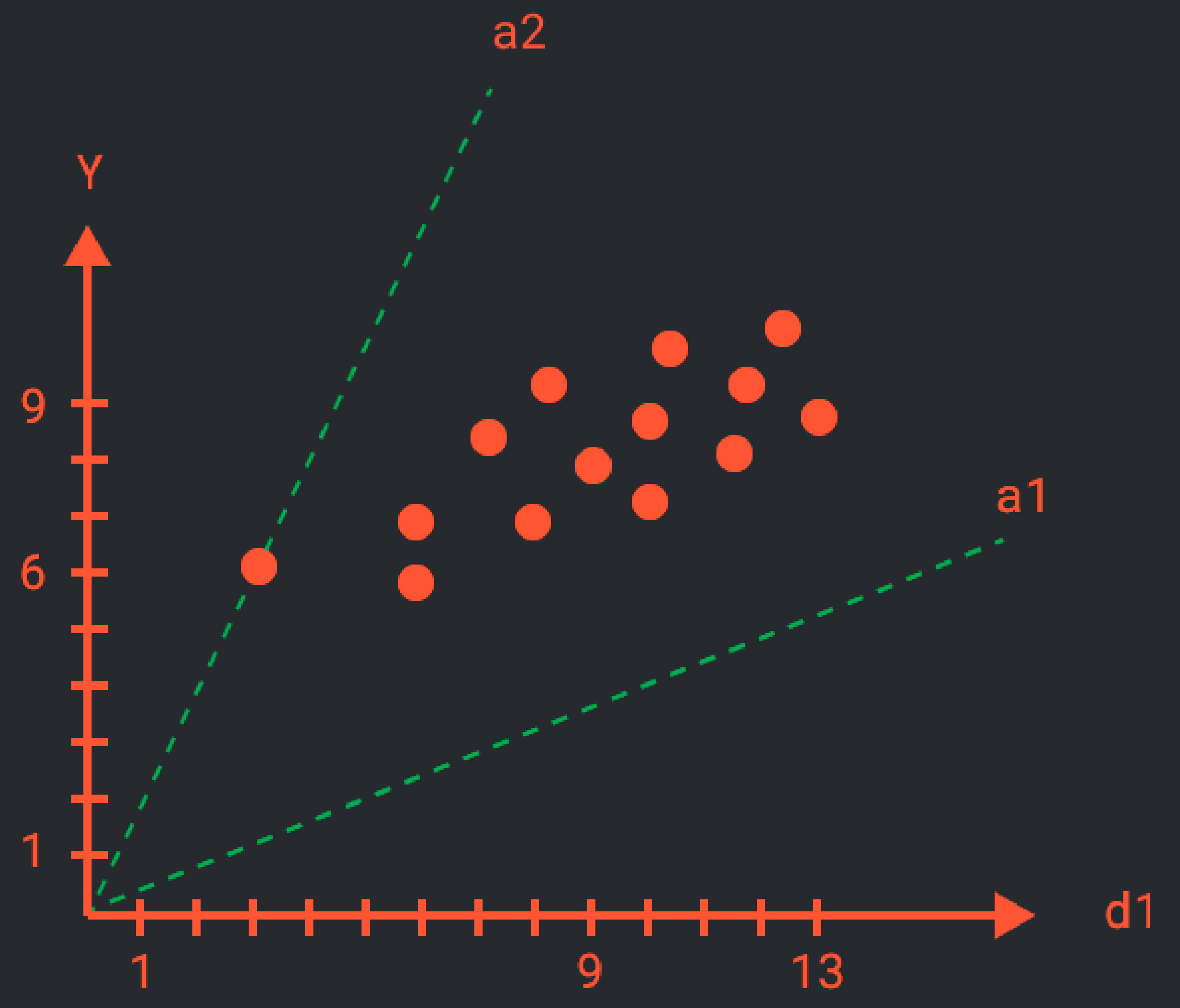


# ЛИНЕЙНЫЕ МОДЕЛИ

Построим нашу первую линейную модель!

$$a(x) = \beta_1 \cdot d_1$$

$$a_1(x) = 0.5 \cdot d_1$$
$$a_2(x) = 2 \cdot d_1$$

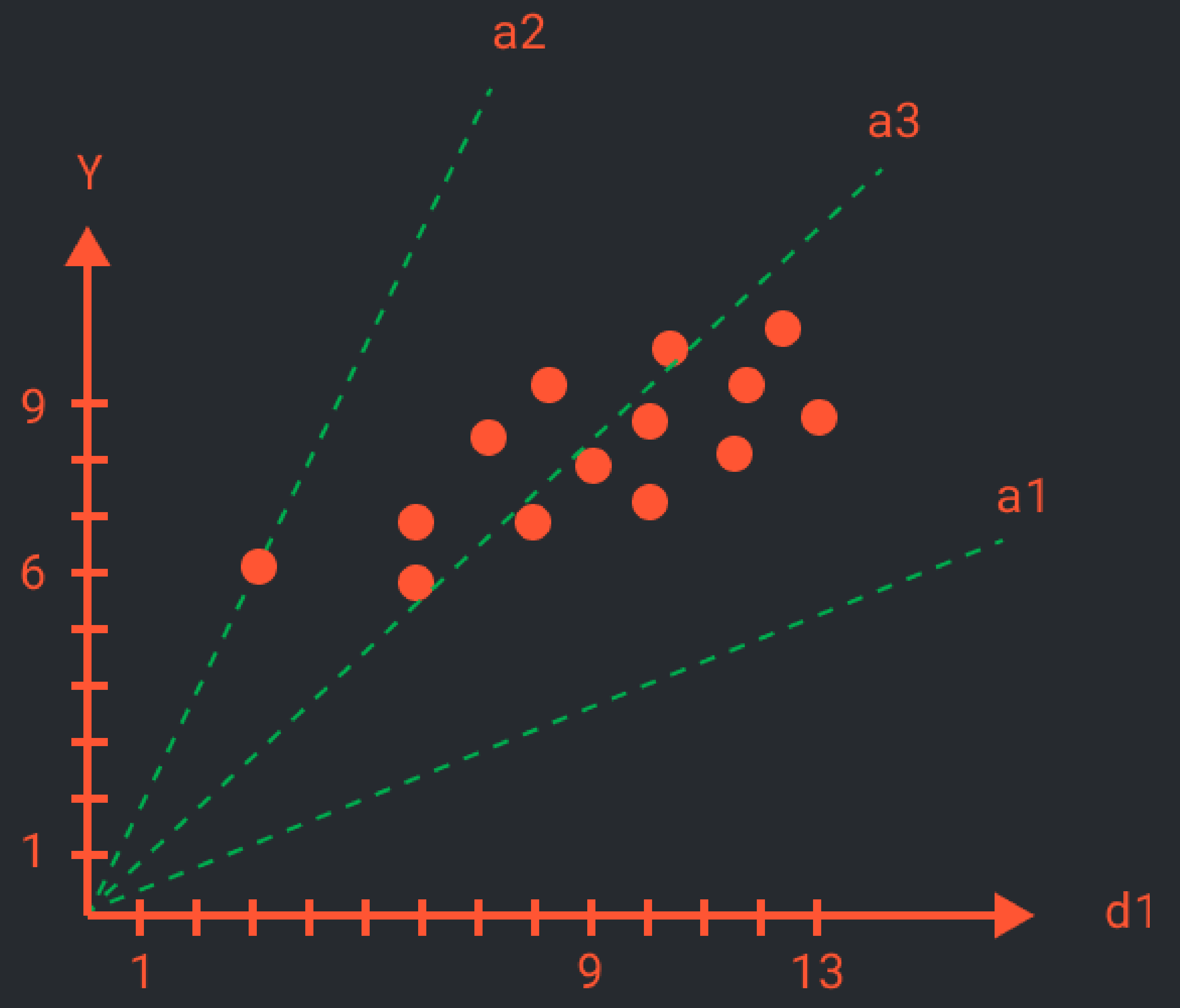


# ЛИНЕЙНЫЕ МОДЕЛИ

Построим нашу первую линейную модель!

$$a(x) = \beta_1 \cdot d_1$$

$$a_1(x) = 0.5 \cdot d_1$$
$$a_2(x) = 2 \cdot d_1$$
$$a_3(x) = 1 \cdot d_1$$




Это целое семейство моделей!

Линейных моделей.

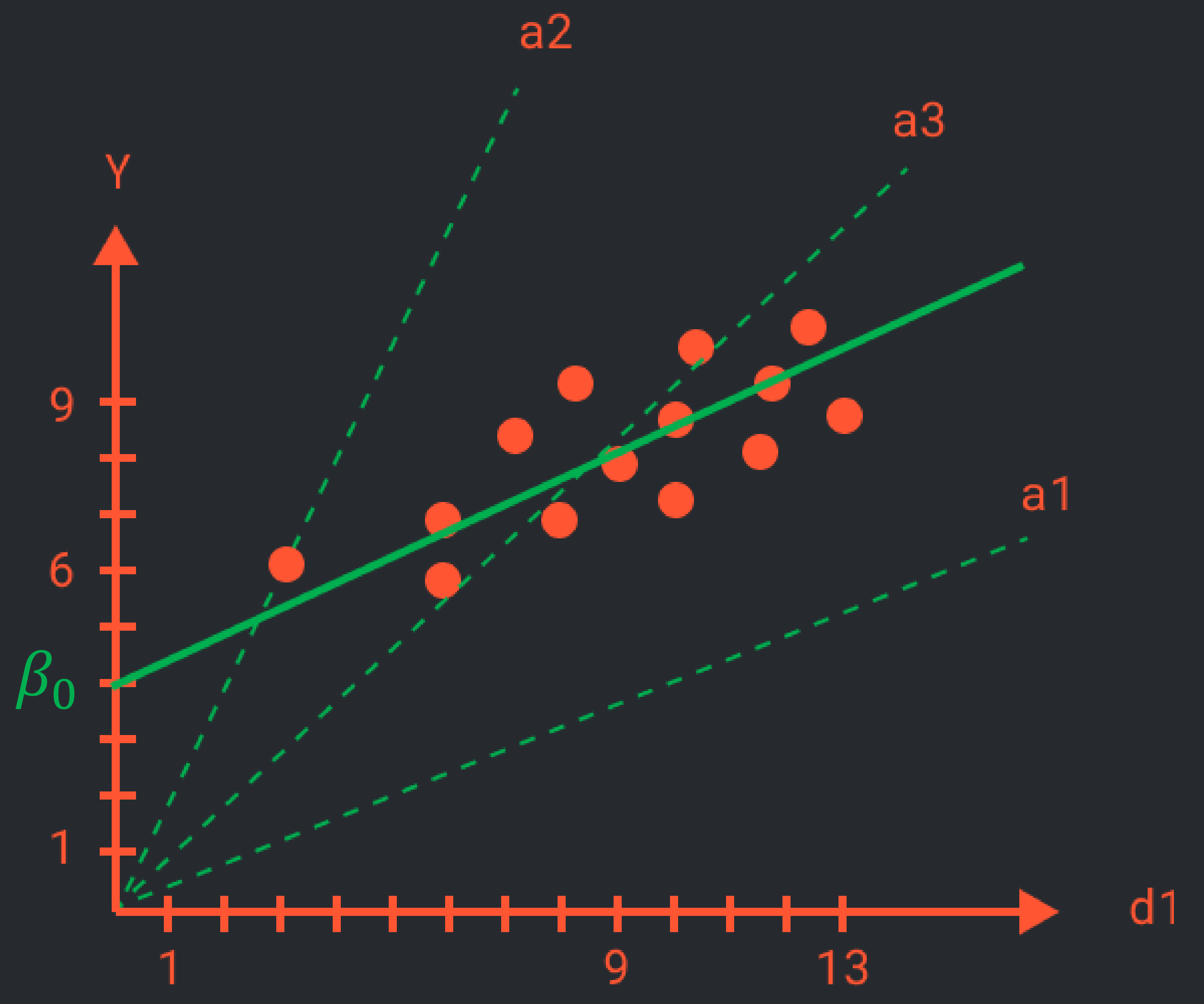
# ЛИНЕЙНЫЕ МОДЕЛИ

Обычно к линейному семейству принято добавлять свободный коэффициент

$$a(x) = \beta_1 \cdot d_1 + \beta_0$$

Это равносильно добавлению константного признака в объекты выборки

$$x_1 = (d_1^1, d_0) = (3, 1) \qquad y_1 = 6$$
$$x_2 = (d_1^2, d_0) = (9, 1) \qquad y_2 = 8$$
$$\dots$$
$$x_n = (d_1^n, d_0) = (13, 1) \qquad y_n = 9$$



# РЕЗЮМЕ

- Узнали, какие модели называют линейными (семейство)
- Чтобы получить конкретный алгоритм, следует подбирать параметры модели
- Стоит добавлять в линейные модели константный признак!

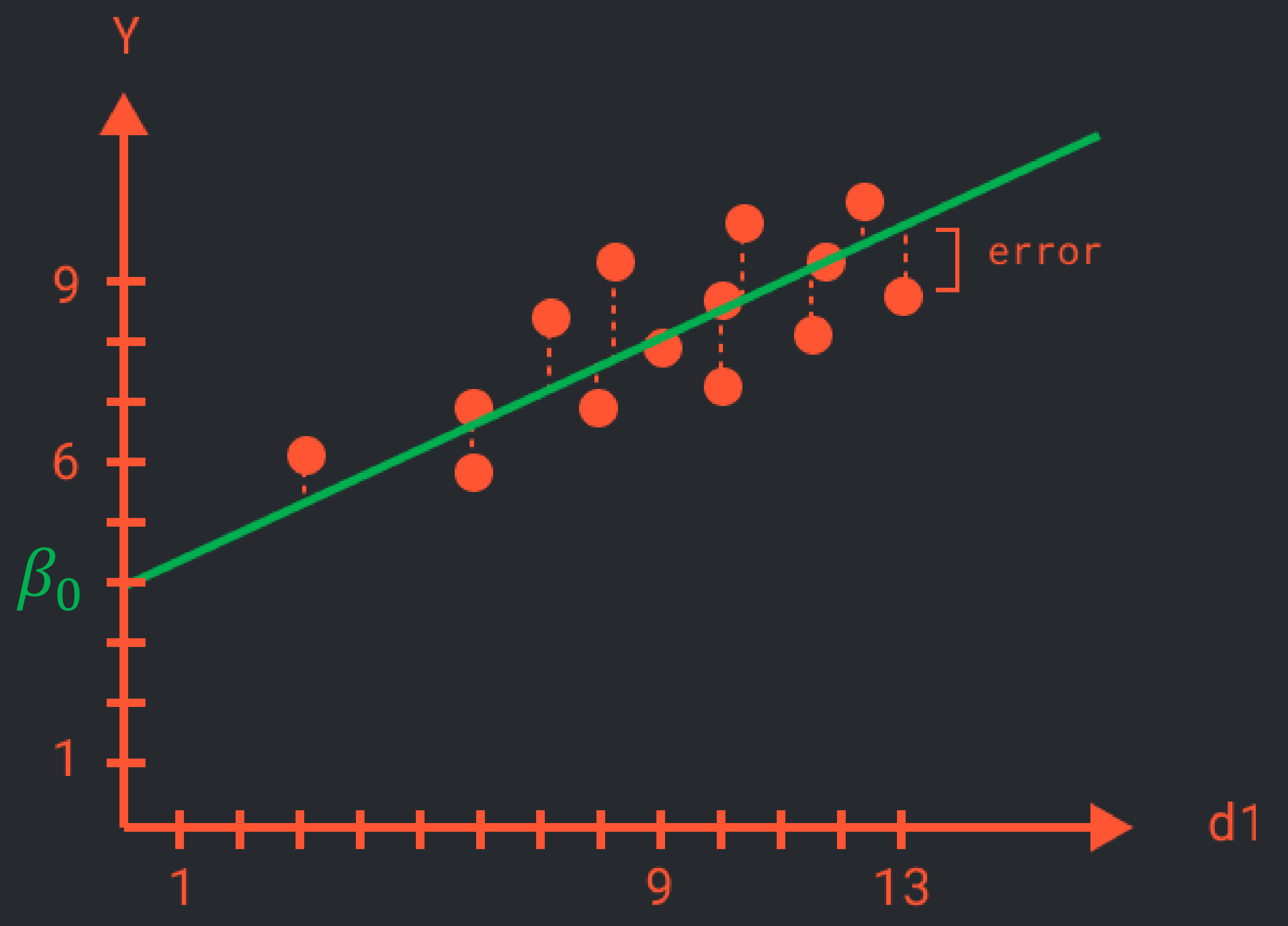
# ЛИНЕЙНЫЕ МОДЕЛИ

## МОДЕЛЬ

Обычно задача МО – это найти такие параметры в выбранном семействе, при которых зависимость между $Y$ и $X$ оценена наиболее правдоподобно.

При правильно выбранном Loss'е, это равносильно поиску минимума функционала качества.

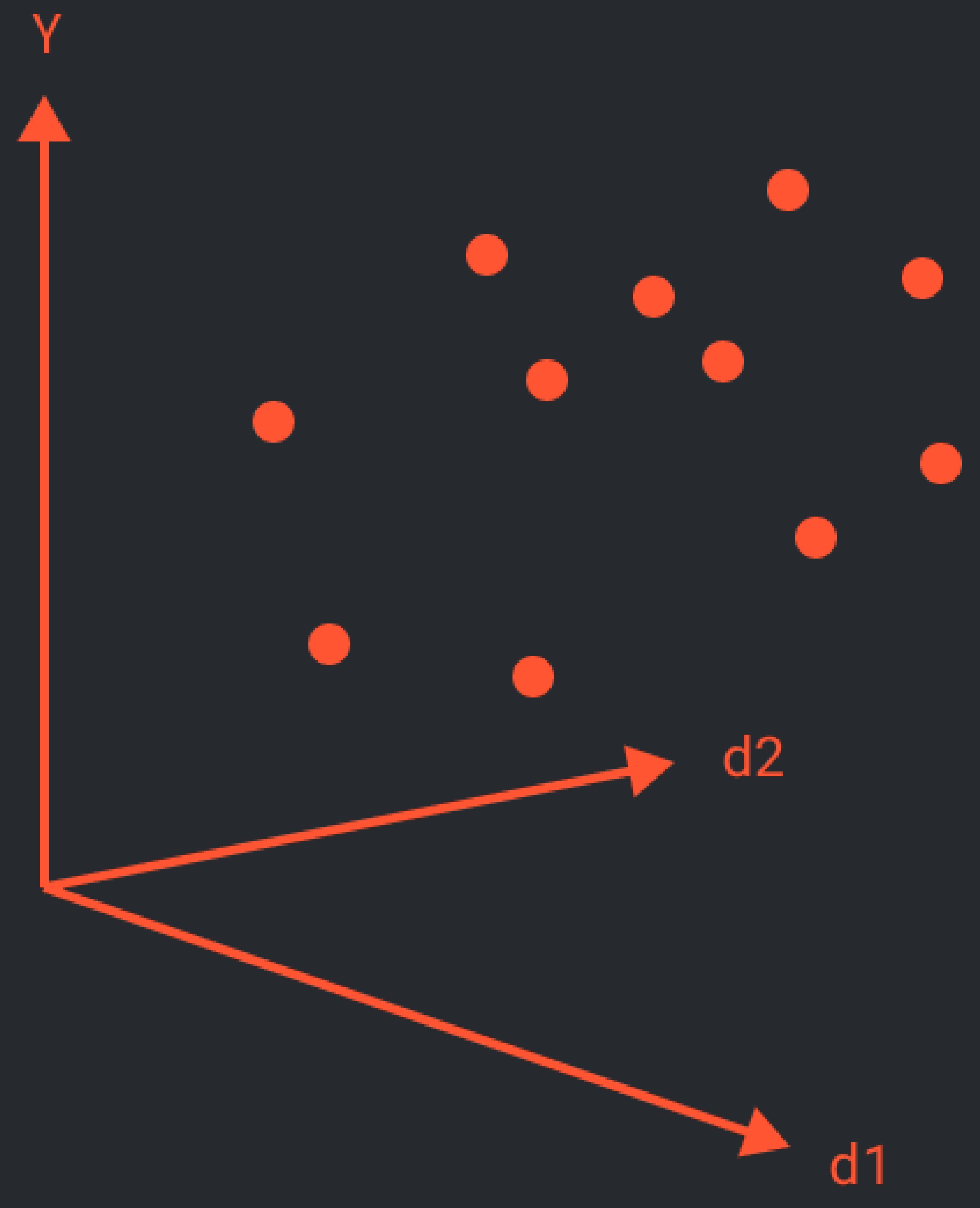
$$\beta^* = argmin\, Q(a(x), X)$$



# МНОГОМЕРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Теперь имеем набор объектов с 2 признаками

$$x_1 = (d_1^1, d_2^1) = (3, 2) \qquad y_1 = 6$$
$$x_2 = (d_1^2, d_2^2) = (9, 5) \qquad y_2 = 8$$
$$\dots$$
$$x_n = (d_1^n, d_2^n) = (13, 9) \qquad y_n = 9$$

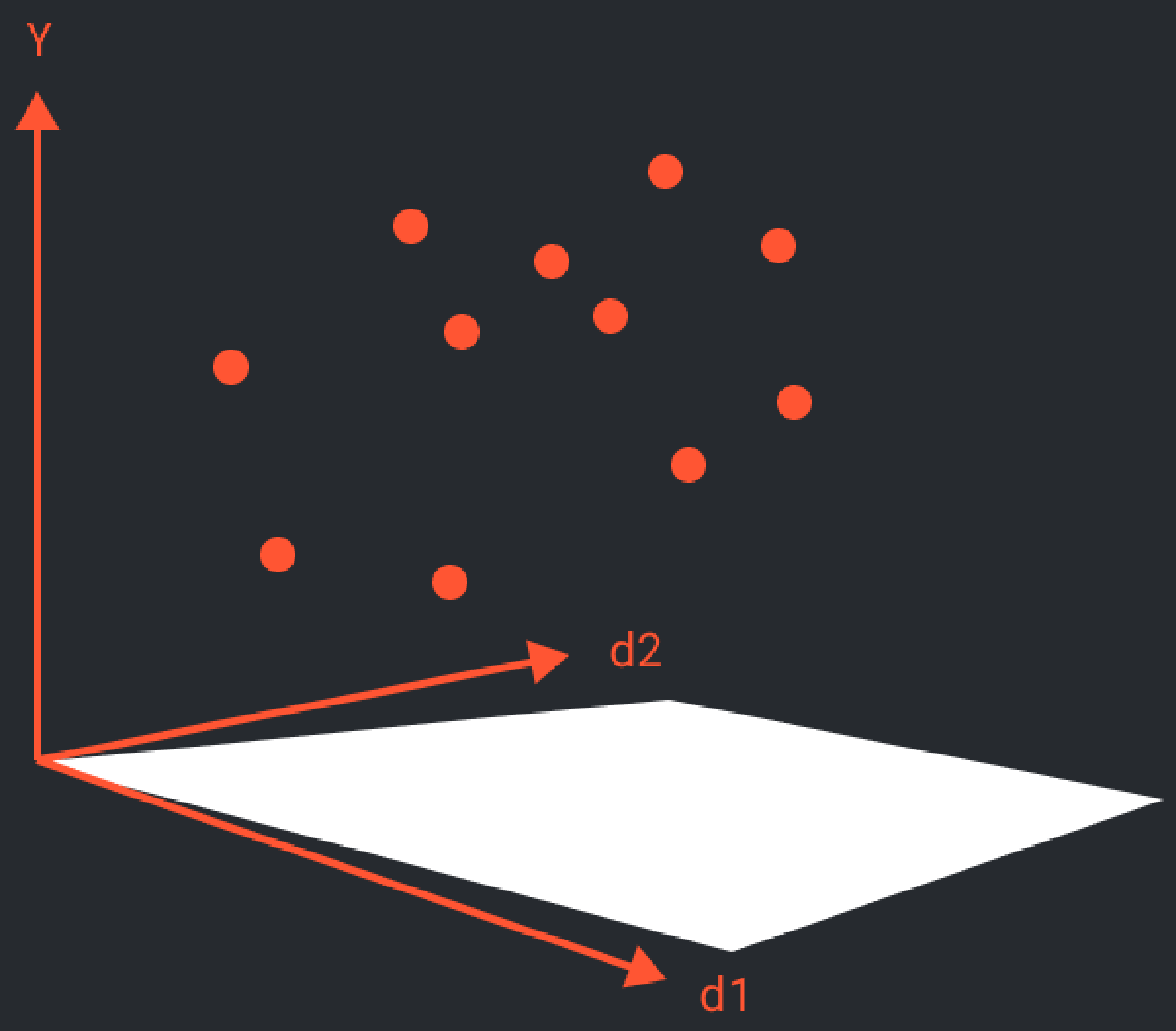


# МНОГОМЕРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Теперь строим не прямую, а плоскость!

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

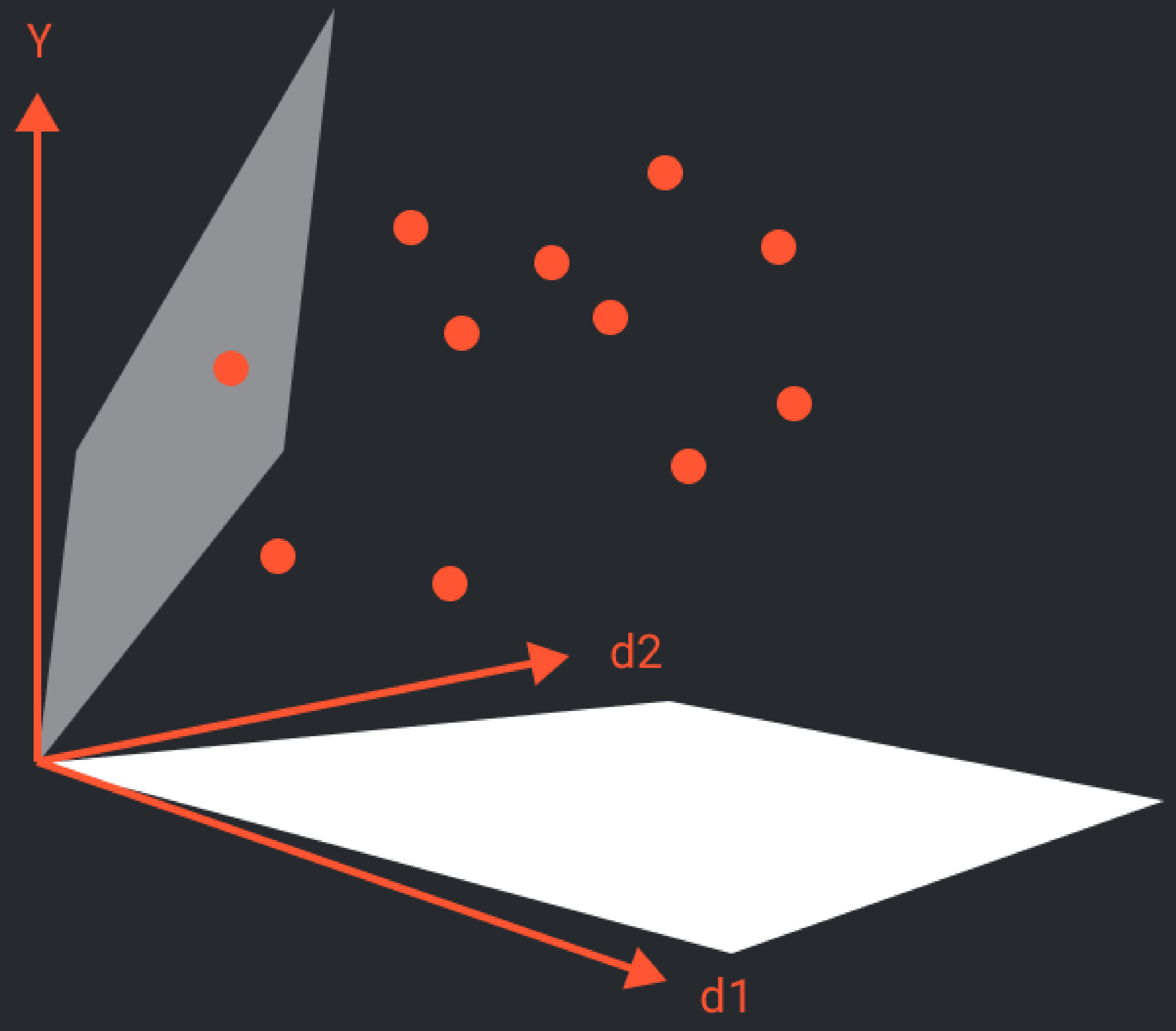
$$a_1(x) = 0 \cdot d_1 + 0.2 \cdot d_2$$



# МНОГОМЕРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Теперь строим не прямую, а плоскость!

$$a(x) = \beta_1 \cdot d_1 + \beta_2 \cdot d_2$$
$$a_1(x) = 0 \cdot d_1 + 0.2 \cdot d_2$$
$$a_2(x) = 5 \cdot d_1 + 4 \cdot d_2$$



# МНОГОМЕРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

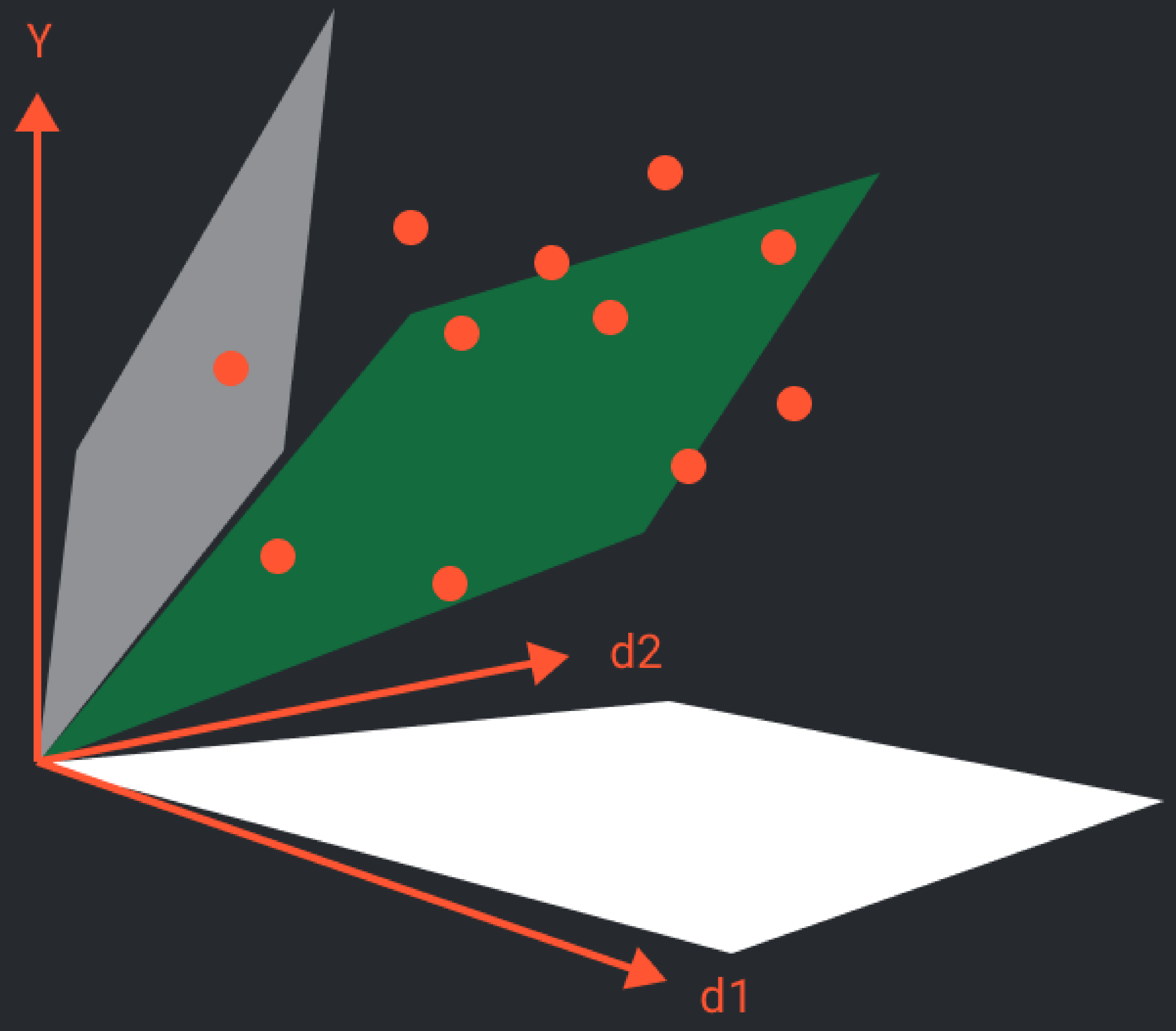
Теперь строим не прямую, а плоскость!

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

$$a_1(x) = 0 \cdot d_1 + 0.2 \cdot d_2$$

$$a_2(x) = 5 \cdot d_1 + 4 \cdot d_2$$

$$a_3(x) = 2 \cdot d_1 + 3 \cdot d_2$$




Но еще лучше, если добавим константу!

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

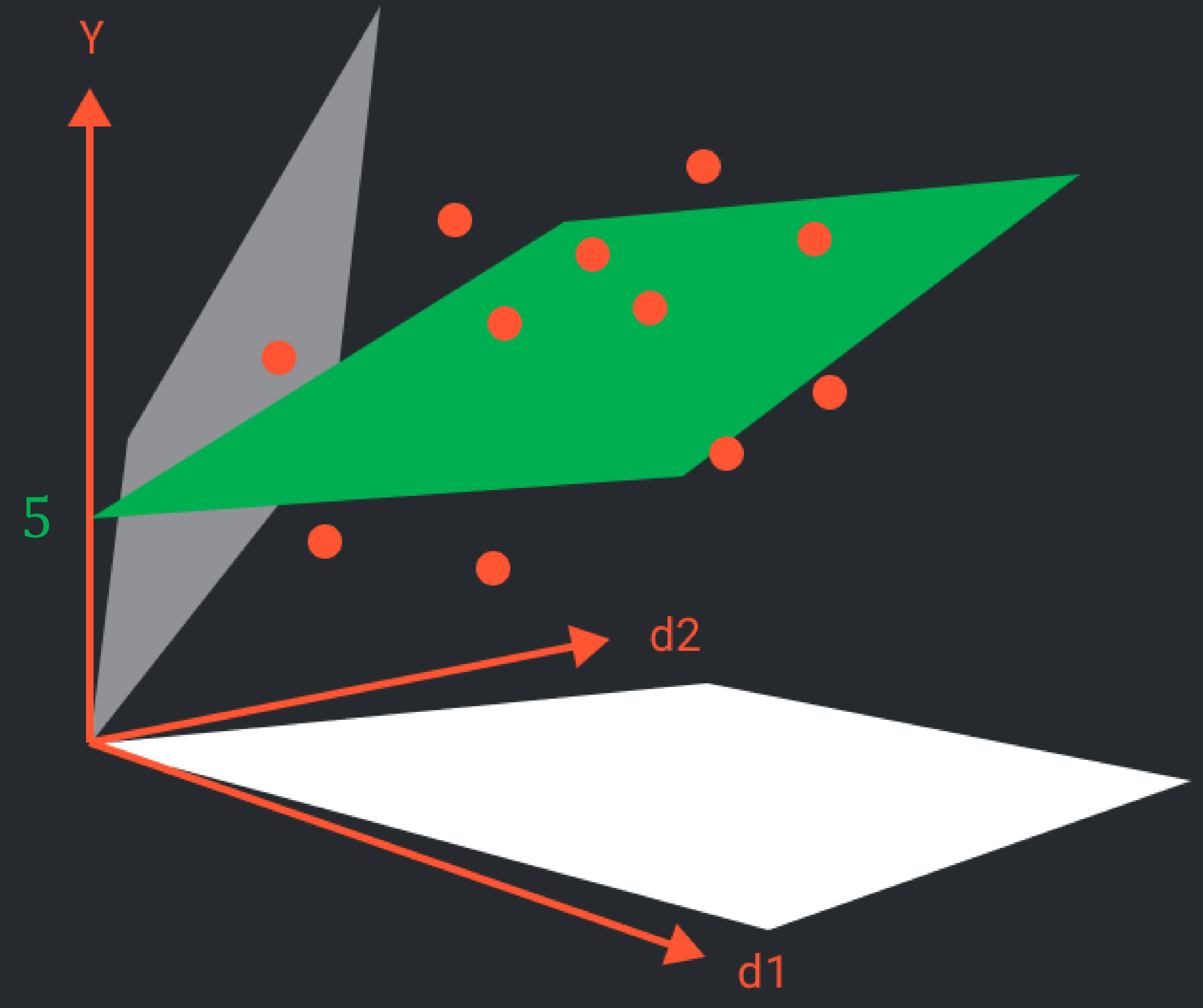
# МНОГОМЕРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

Теперь строим не прямую, а плоскость!

$$a(x) = \beta_1 \cdot d_1 + \beta_2 \cdot d_2$$
$$a_1(x) = 0 \cdot d_1 + 0.2 \cdot d_2$$
$$a_2(x) = 5 \cdot d_1 + 4 \cdot d_2$$

Но еще лучше, если добавим константу!

$$a(x) = \beta_1 \cdot d_1 + \beta_2 \cdot d_2 + \beta_0$$
$$a_3(x) = 2 \cdot d_1 + 3 \cdot d_2 + 5$$



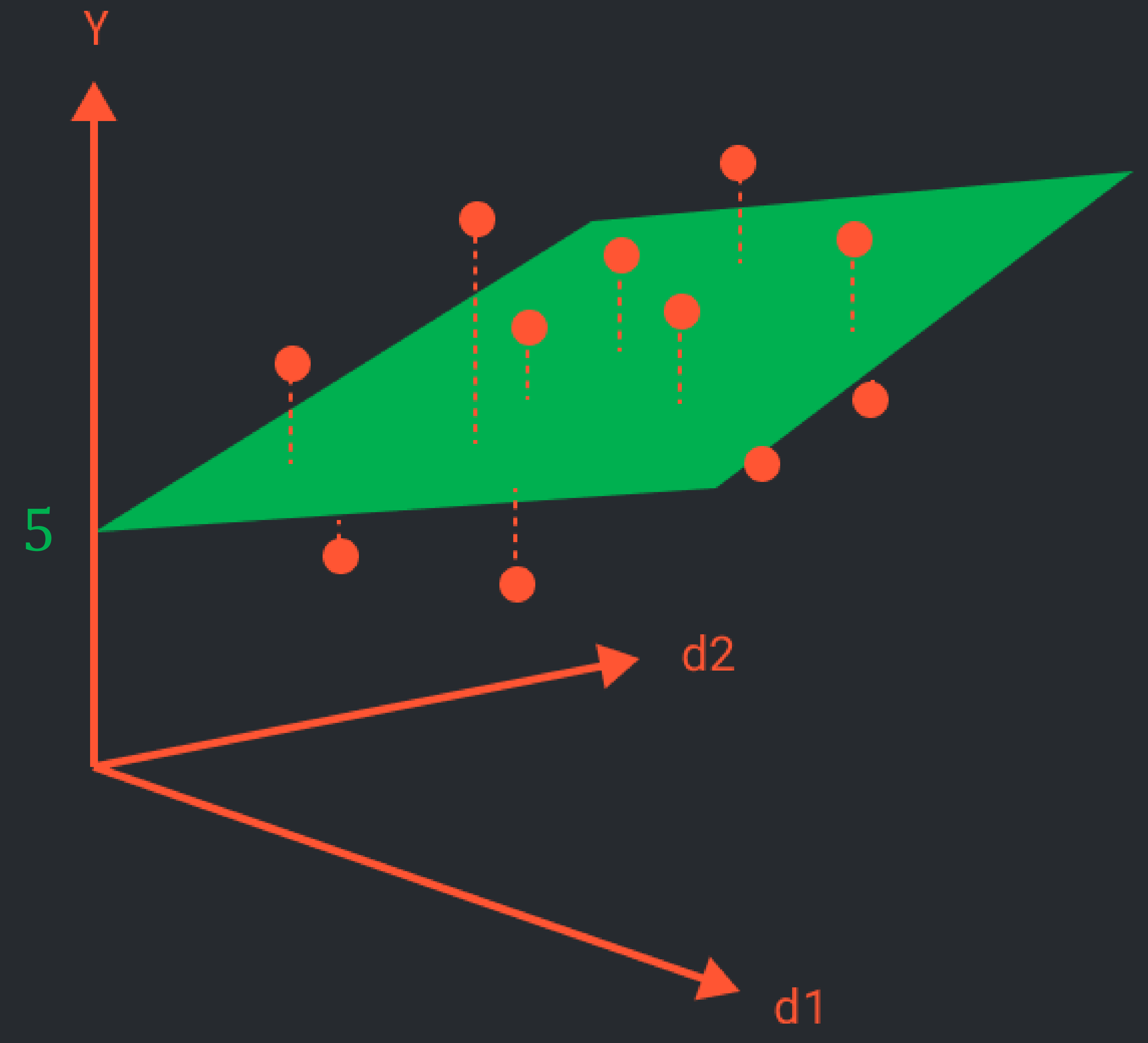
# МНОГОМЕРНАЯ ЛИНЕЙНАЯ РЕГРЕССИЯ

$$a(x) = \beta_1 \cdot d_1 + \beta_2 \cdot d_2 + \beta_0$$
$$a_3(x) = 2 \cdot d_1 + 3 \cdot d_2 + 5$$

$\beta_1^* = 2 \qquad \beta_2^* = 3 \qquad \beta_0^* = 5$

Лучшие параметры модели – те, которые минимизируют функционал ошибки!

$$\beta^* = argmin\, Q(a(x), X)$$



# ЛИНЕЙНАЯ РЕГРЕССИЯ OLS: ПРИМЕР

Найти лучшие $\beta$ значит найти такие параметры, при которых $Q(a(x), X)$ минимальна.

Пусть пытаемся предсказать цену акции компаний в $ через год по текущей цене $(d^1)$ и сумме кредиторской задолженности в миллионах $ $(d^2)$. Допустим, в выборку попало всего 3 компании $(x_1, x_2, x_3)$ с соответствующими ответами $(y_1, y_2, y_3)$.

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

# ЛИНЕЙНАЯ РЕГРЕССИЯ OLS: ПРИМЕР

Пока мы не знаем лучшие коэффициенты нашей линейной модели, но можем в общем виде для каждого объекта выписать нашу оценку в зависимости от $\beta$

Общее уравнение: $a(x) = \beta_1 \cdot d_1 + \beta_2 \cdot d_2 + \beta_0$

Подставим первый объект: $a(x_1) = \beta_1 \cdot 23 + \beta_2 \cdot 0.5 + \beta_0$

Подставим второй объект: $a(x_2) = \beta_1 \cdot 35 + \beta_2 \cdot 1 + \beta_0$

Подставим третий объект: $a(x_3) = \beta_1 \cdot 18 + \beta_2 \cdot 0 + \beta_0$

# ЛИНЕЙНАЯ РЕГРЕССИЯ OLS: ПРИМЕР

Тогда можем и среднеквадратичное отклонение тоже записать, то есть наш функционал:

$$Q(a(x), X) = \frac{1}{m}\sum_{i}^{m}(a(x_i) - y_i)^2 =$$

$$\frac{1}{3}\cdot[(a(x_1) - y_1)^2 + (a(x_2) - y_2)^2 + (a(x_3) - y_3)^2] =$$

$$\boxed{\begin{array}{c} \frac{1}{3}\cdot[\,(\beta_1 \cdot 23 + \beta_2 \cdot 0.5 + \beta_0 - 55)^2 + \\ (\beta_1 \cdot 35 + \beta_2 \cdot 1 + \beta_0 - 100)^2 + \\ (\beta_1 \cdot 18 + \beta_0 - 45)^2] \end{array}} \rightarrow min$$

# РЕЗЮМЕ

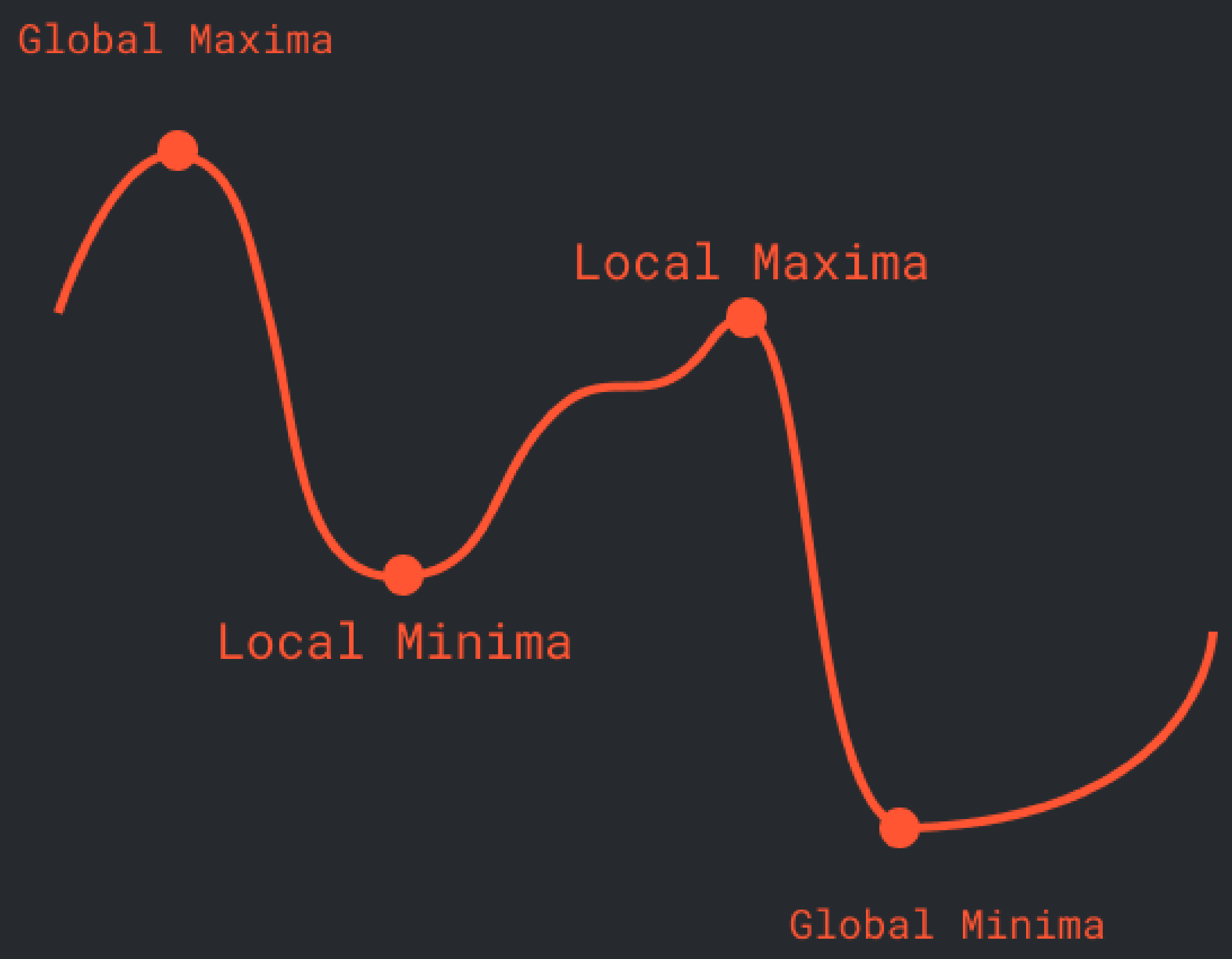
- Научились формально составлять задачу поиска оптимальных параметров регрессии
- Осталось научиться находить минимумы и максимумы математических функций!

# ЛИКБЕЗ №2: ПРОИЗВОДНЫЕ И ЭКСТРЕМУМЫ

## ПОИСК МИНИМУМОВ И МАКСИМУМОВ

Обратим внимание, что экстремумы (ямочки) возникают там, где функция меняет характер роста или убывания.

Круто было бы иметь инструмент, который в каждой точке показывал бы, убывает функция или растет!

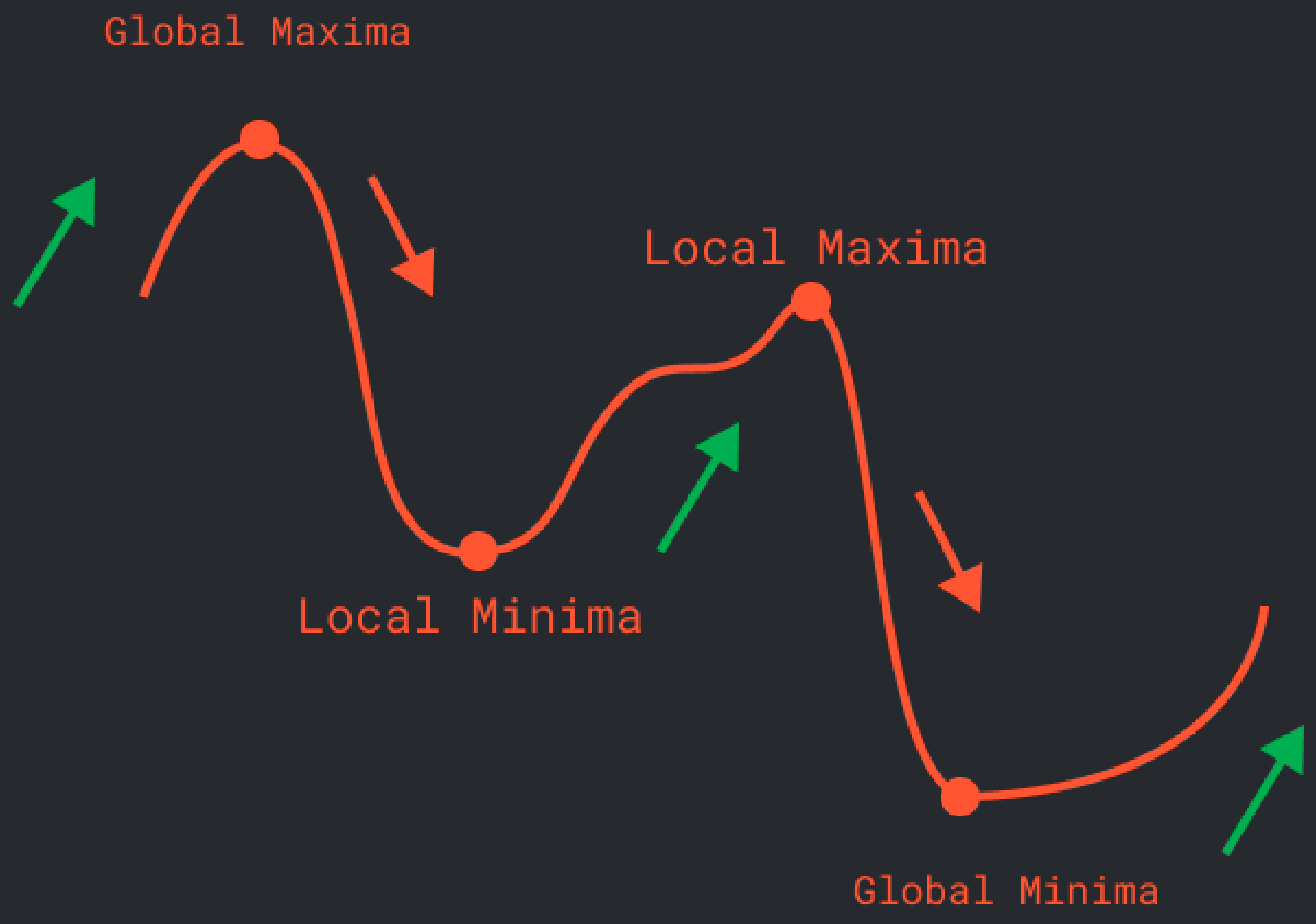


# ЛИКБЕЗ №2: ПРОИЗВОДНЫЕ И ЭКСТРЕМУМЫ

## ПОИСК МИНИМУМОВ И МАКСИМУМОВ

Обратим внимание, что экстремумы (ямочки) возникают там, где функция меняет характер роста или убывания.

Круто было бы иметь инструмент, который в каждой точке показывал бы, убывает функция или растет!

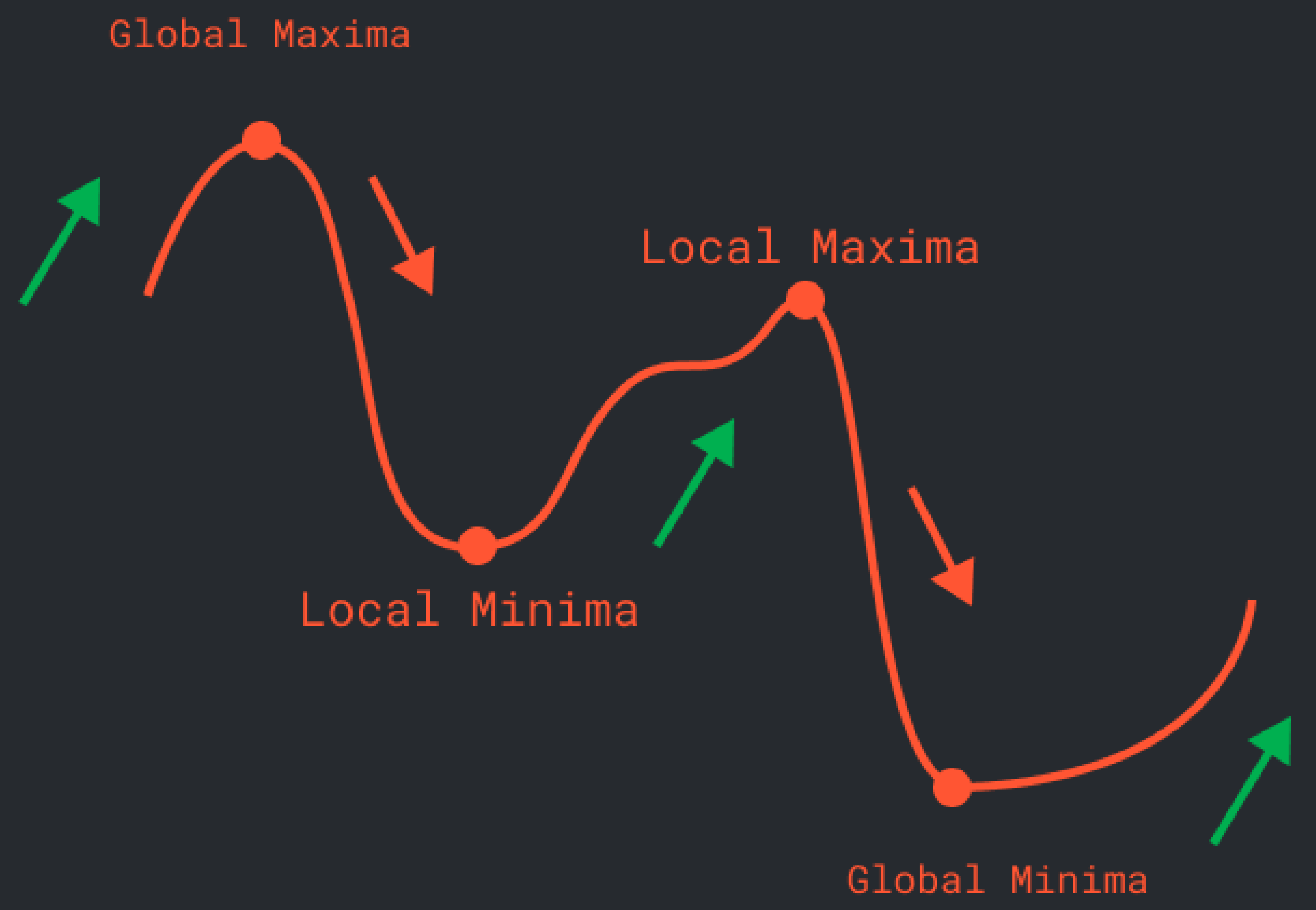


# ЛИКБЕЗ №2: ПРОИЗВОДНЫЕ И ЭКСТРЕМУМЫ

ПРОИЗВОДНАЯ: $y'_x$

Производная показывает, в каком соотношении меняется значение функции, если совсем чуть-чуть изменить значение аргумента:

$$y'_x = \frac{\Delta y}{\Delta x} \qquad \Delta x \approx 0$$



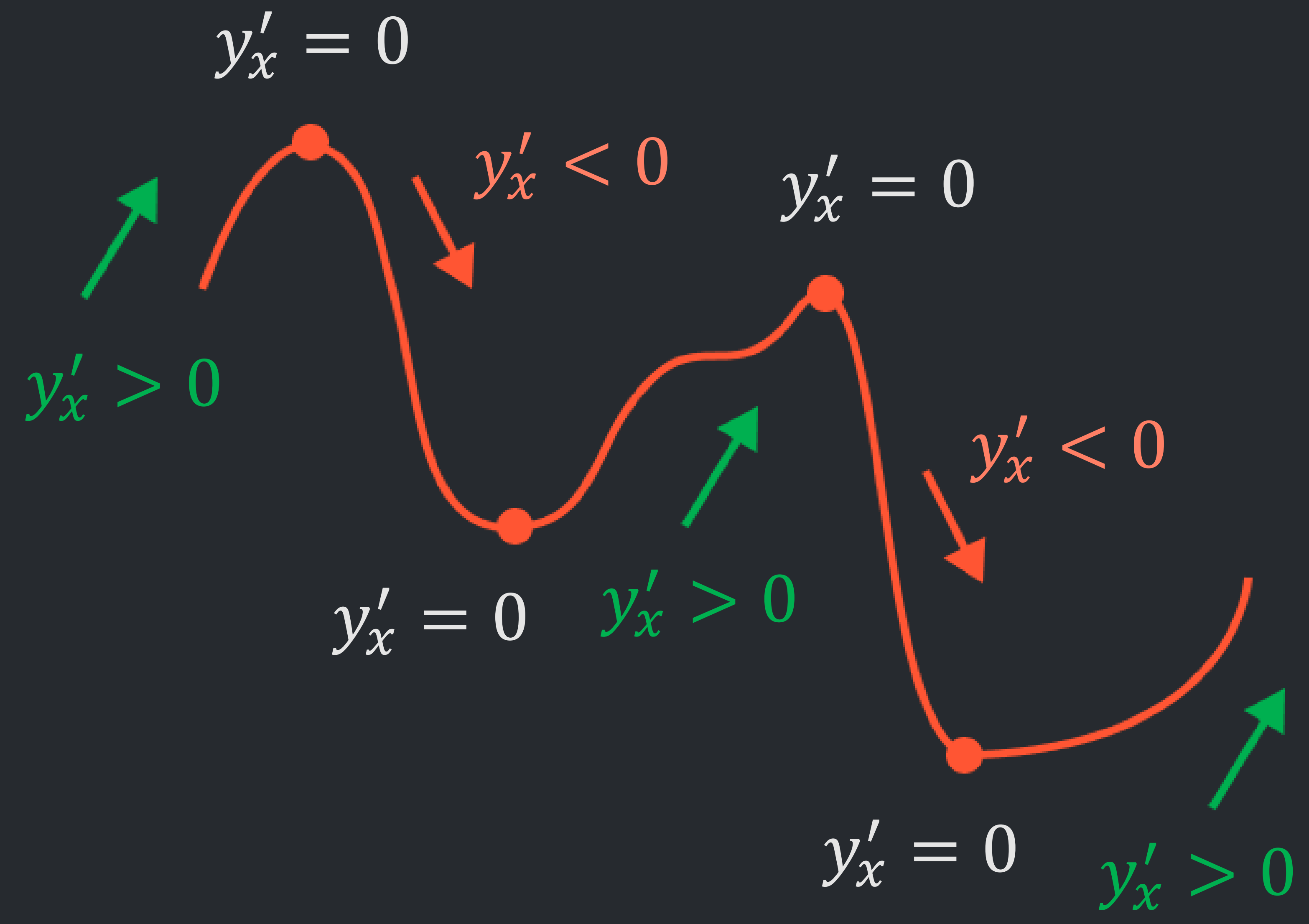
# ЛИКБЕЗ №2: ПРОИЗВОДНЫЕ И ЭКСТРЕМУМЫ

ПРОИЗВОДНАЯ: $y'_x$

$y'_x > 0 \rightarrow$ функция растет!

$y'_x < 0 \rightarrow$ функция убывает!

$y'_x = 0 \rightarrow$ интересная точка!



# ЛИКБЕЗ №2: ПРОИЗВОДНЫЕ И ЭКСТРЕМУМЫ

## КАК СЧИТАТЬ ПРОИЗВОДНЫЕ?

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

$$\frac{\Delta y}{\Delta x} = \frac{\mathrm{y}(\mathrm{x} + \Delta x) - y(x)}{\Delta x}$$

$$P.S. \quad \Delta x \to 0$$

Пусть $y(x) = x^2$

$$y'_x = \frac{(\mathrm{x} + \Delta x)^2 - x^2}{\Delta x} = \frac{x^2 + 2 \cdot \mathrm{x} \cdot \Delta x + (\Delta x)^2 - x^2}{\Delta x} =$$

$$\frac{2 \cdot x \cdot \Delta x + (\Delta x)^2}{\Delta x} = \frac{2 \cdot x \cdot \Delta x}{\Delta x} + \frac{(\Delta x)^2}{\Delta x} = 2 \cdot x + \Delta x = 2 \cdot x$$

# ЛИКБЕЗ №2: ПРОИЗВОДНЫЕ И ЭКСТРЕМУМЫ

## ТАБЛИЦА ПРОИЗВОДНЫХ

$(x^n)' = n \cdot x^{n-1}$

$(a^x)' = a^x \cdot \ln(a)$

$(\log_a x)' = \frac{1}{x \cdot \ln a}$

$(\sin x)' = \cos x$

$(\sin x)' = \cos x$

$(\cos x)' = -\sin x$

$(\operatorname{tg} x)' = \frac{1}{\cos^2 x}$

$(\operatorname{ctg} x)' = -\frac{1}{\sin^2 x}$

$(\arcsin x)' = \frac{1}{\sqrt{1-x^2}}$

$(\arccos x)' = \frac{-1}{\sqrt{1-x^2}}$

$(\operatorname{arctg} x)' = \frac{1}{1+x^2}$

$(\operatorname{arcctg} x)' = \frac{-1}{1+x^2}$

# ЛИКБЕЗ №2: ПРОИЗВОДНЫЕ И ЭКСТРЕМУМЫ

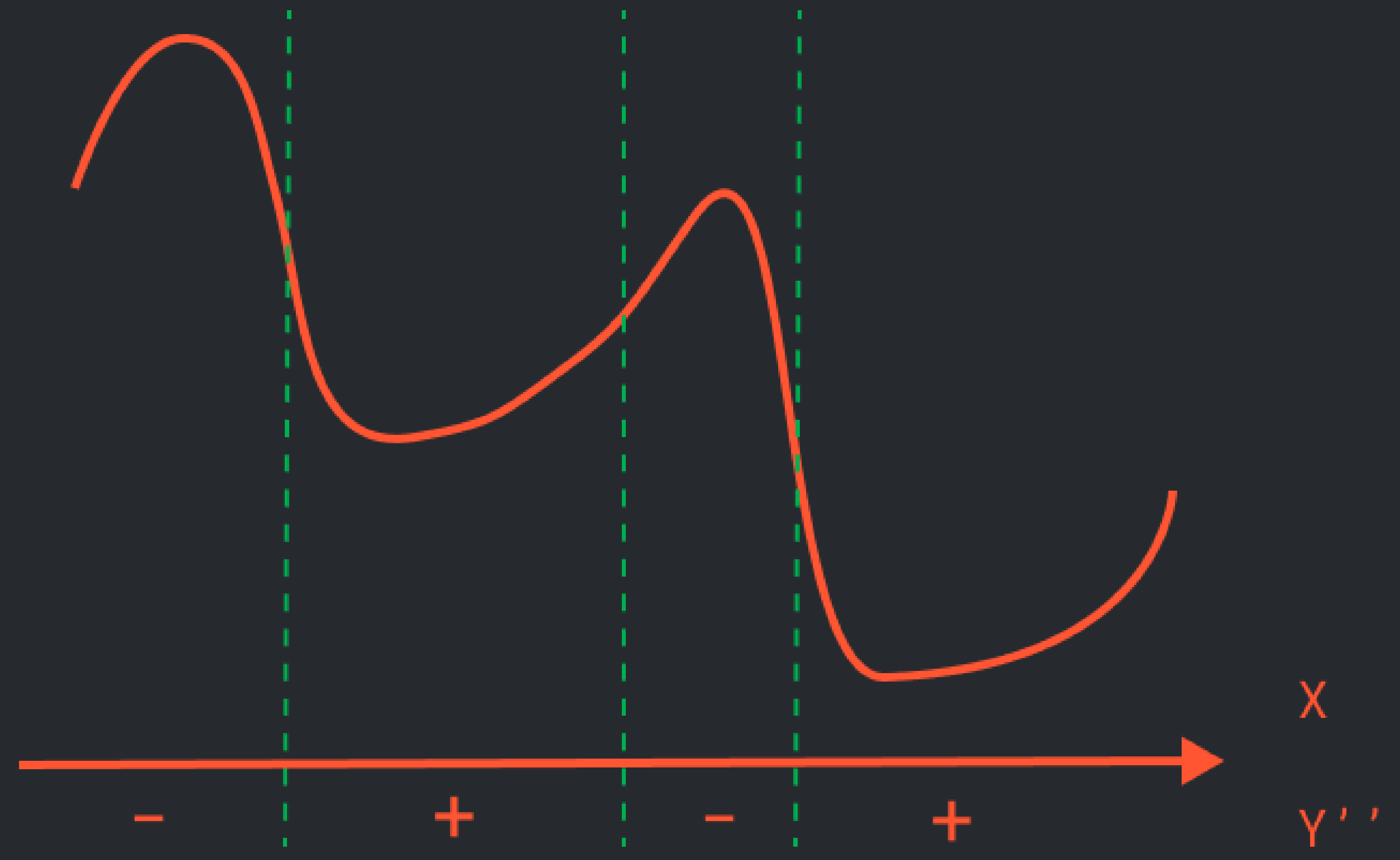
## УСЛОВИЕ ВТОРОГО ПОРЯДКА

Заметим, чтобы различить минимумы и максимумы между собой, должно выполняться соответственно условие выпуклости / вогнутости – за это отвечает вторая производная.

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

$$y'' = y' > 0 \rightarrow \text{функция выпуклая}$$

$$y'' = y' < 0 \rightarrow \text{функция вогнутая}$$



# ЛИКБЕЗ №2: ПРОИЗВОДНЫЕ И ЭКСТРЕМУМЫ

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

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

$$(x^3)' - 6 \cdot (x^2)' - 15 \cdot (x)' + (10)' =$$

$$3x^2 - 12x - 15 = 0$$

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


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

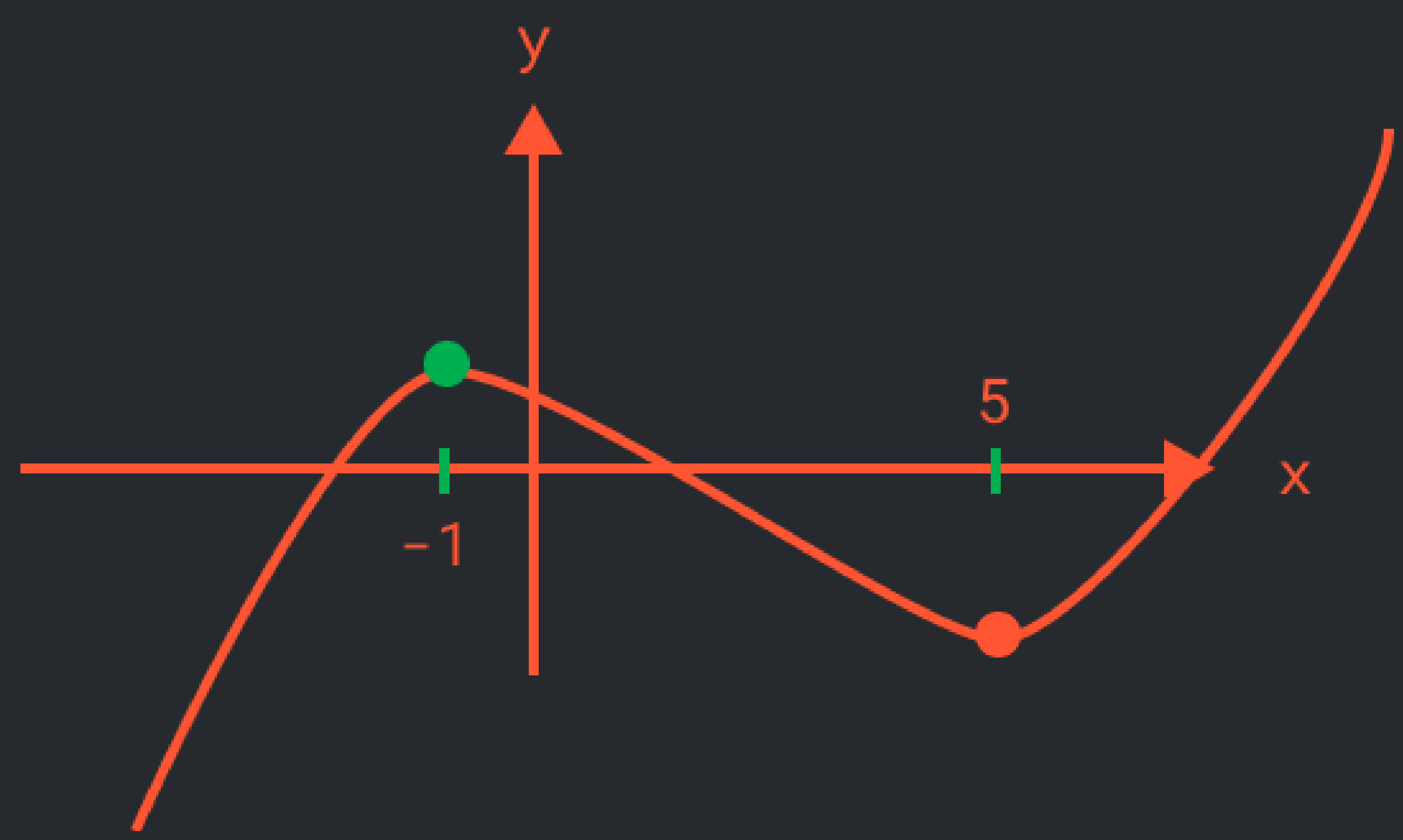
$$y'(0) = 0 - 0 - 15 = -15 < 0$$

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

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

$y''(-1) = -18 < 0 \Rightarrow$ это локальный максимум

$y''(5) = +18 > 0 \Rightarrow$ это локальный минимум



# ЛИКБЕЗ №2: ПРОИЗВОДНЫЕ И ЭКСТРЕМУМЫ

## ФУНКЦИИ НЕСКОЛЬКИХ ПЕРЕМЕННЫХ

Логика та же самая – чтобы найти критические точки, приравниваем первые производные к 0, решаем систему.

$$z(x;y) = x^2 + 3y^2$$

$$\begin{cases} z'_x = 2 \cdot x = 0 \\ z'_y = 6 \cdot y = 0 \end{cases} \rightarrow \quad (x^*, y^*) = (0;0)$$

## УСЛОВИЕ ВТОРОГО ПОРЯДКА

Чтобы проверить выпусклость / вогнутость в окрестности критической точки, применяем критерий Сильвестра на матрицу из частных производных 2 порядка.

$$z''_{xx} = (z'_x)'_x = 2 \qquad z''_{xy} = (z'_x)'_y = 0$$

$$z''_{yx} = \left(z'_y\right)'_x = 0 \qquad z''_{yy} = \left(z'_y\right)'_y = 6$$

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

# РЕЗЮМЕ

- Выяснили, что такое производная и о чем она сигнализирует
- Научились выводить и находить производные различных функций
- На примере посмотрели, как можно исследовать функцию на экстремумы
- Узнали про условие второго порядка
- Масштабировали знания на функции нескольких переменных

# ЛИНЕЙНАЯ РЕГРЕССИЯ OLS: ПРИМЕР

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

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

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

Выпуклость очевидна из формы функционала. Но можно и вправду проверить еще и матрицу Гессе.

Итого, лучшая линейная модель для нашей выборки:

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

# ЛИНЕЙНАЯ РЕГРЕССИЯ OLS: ПРИМЕР

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

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

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

Выпуклость очевидна из формы функционала. Но можно и вправду проверить еще и матрицу Гессе.

Итого, лучшая линейная модель для нашей выборки:

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

# ЛИНЕЙНАЯ РЕГРЕССИЯ OLS: ПРИМЕР

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

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

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

# ЛИКБЕЗ №3: МАТРИЦЫ

Матрицы – форма хранения чисел в математике! Бывают разных размеров.
Матрицы с 1 строкой или 1 столбцом часто называют векторами.

$$\begin{pmatrix} 23 & 0{,}5 \\ 35 & 1 \\ 18 & 0 \end{pmatrix}$$

$3\ x\ 2$

$$\begin{pmatrix} 55 \\ 100 \\ 45 \end{pmatrix}$$

$3\ x\ 1$

# ЛИКБЕЗ №3: МАТРИЧНОЕ СЛОЖЕНИЕ / ВЫЧИТАНИЕ

Складывать и вычитать можно ИСКЛЮЧИТЕЛЬНО матрицы одного размера!

Делается это поэлементно!

$$\underset{3\ x\ 2}{\begin{pmatrix} 23 & 0,5 \\ 35 & 1 \\ 18 & 0 \end{pmatrix}} + \underset{3\ x\ 2}{\begin{pmatrix} 30 & 0 \\ 40 & 3 \\ 10 & 0 \end{pmatrix}} = \underset{3\ x\ 2}{\begin{pmatrix} 53 & 0,5 \\ 75 & 4 \\ 28 & 0 \end{pmatrix}}$$

# ЛИКБЕЗ №3: МАТРИЧНОЕ УМНОЖЕНИЕ

Определено матричное умножение $A \cdot B$, если $A$ – размера $n \, x \, k$, $B$ – размера $k \, x \, m$. Причем тогда результат умножения этих двух матриц – матрица C – размера $n \, x \, m$. Причем чтобы получить $c_{ij}$, нужно скалярно перемножить $i$ строку $A$ и $j$ столбец $B$. Например,

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

# ЛИКБЕЗ №3: ТРАНСПОНИРОВАНИЕ МАТРИЦЫ

Транспонировать матрицу – повернуть ее на бок, т.е. столбцы превратить в строки и наоборот.

$$A = \begin{pmatrix} 23 & 0{,}5 \\ 35 & 1 \\ 18 & 0 \end{pmatrix} \xrightarrow{T} A^{\mathrm{T}} = \begin{pmatrix} 23 & 35 & 18 \\ 0{,}5 & 1 & 0 \end{pmatrix}$$

$3\ x\ 2$ $\qquad$ $2\ x\ 3$

# ЛИКБЕЗ №3: ЕДИНИЧНАЯ МАТРИЦА

Ее принято обозначать символом $E$. Она может быть любого квадратного размера. Аналог 1.

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

# ЛИКБЕЗ №3: ОБРАТНЫЕ МАТРИЦЫ

Обратная матрица – это некоторый многомерный аналог обратному числу. Точнее, говорят, что у матрицы $A$ существует единственная обратная к ней матрица $A^{-1}$, если

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

$$\overset{A}{\begin{pmatrix} 2 & 5 \\ 3 & 8 \end{pmatrix}} \overset{A^{-1}}{\begin{pmatrix} 8 & -5 \\ -3 & 2 \end{pmatrix}} = \overset{E}{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}$$

# ЛИНЕЙНАЯ РЕГРЕССИЯ OLS: МАТРИЧНАЯ ФОРМА

Вспомним, что всю нашу выборку, состоящую из ряда объектов и признаков можно представить в виде матрицы $X$ размера n x k, где n – количество элементов в выборке, а k – количество признаков (включая константу). Введем вектор коэффициентов $B$ и заметим, что:

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

Проверим это на нашем предыдущем примере

# ЛИНЕЙНАЯ РЕГРЕССИЯ OLS: МАТРИЧНАЯ ФОРМА

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

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

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

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

# ЛИНЕЙНАЯ РЕГРЕССИЯ OLS: МАТРИЧНАЯ ФОРМА

Тогда задача поиска оптимальных коэффициентов сводится к минимизации произведения двух матриц. Можно взять матричный дифференциал и приравнять его к нулю. Этот шаг аналогичен нахождению частных производных и решению системы для поиска критических точек.

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

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

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

# РЕЗЮМЕ

- Изучили матричный аппарат
- Нашли на примере оптимальные коэффициенты регрессии
- Перевели задачу минимизации среднеквадратической ошибки на язык матриц
- Пора практиковаться!

# СПАСИБО

ТАБАКАЕВ НИКИТА