Линейная регрессия

Skillbox

образовательная платформа

Линейная модель регрессии

Множество объектов X

n объектов
k признаков

Каждый из объектов характеризуется вектором признаков x = (x1, x2, ... , xk)

Множество объектов Y

Множество рациональных чисел $Y = \mathbb{R}$

Существует некоторая целевая функция, как зависимость $\hat{y}: X \rightarrow Y$

Задача машинного обучения

Построить такой алгоритм $a: X \to Y$, который по вектору признаков приближал бы целевую функцию $\hat{y}$.

Пример: предсказание стоимости автомобиля

Множество объектов X

n различных автомобилей
k признаков:

класс автомобиля, марка авто, объём двигателя, новая машина или б/у и т. д.

Множество объектов Y

Стоимость автомобиля

Существует некоторая целевая функция,
как зависимость цены на автомобиль от его признаков
$\hat{y}: X \to Y$

Уравнение регрессии

$$\hat{y} = f(w, x) = w_0 + w_1 \times x_1 + \ldots + w_k \times x_k$$

Где

$\hat{y}$ — целевая переменная,

$(x_1, \ldots, x_k)$ — вектор признаков,

$w_0, w_1, \ldots, w_k$ — параметры модели цели,

как его ещё называют

$w_1, \ldots, w_k$ — вектор весов,

а число w_0 — свободный коэффициент, или сдвиг (bias)

Или компактная запись

$$\hat{y} = \langle x, w \rangle + w_0$$

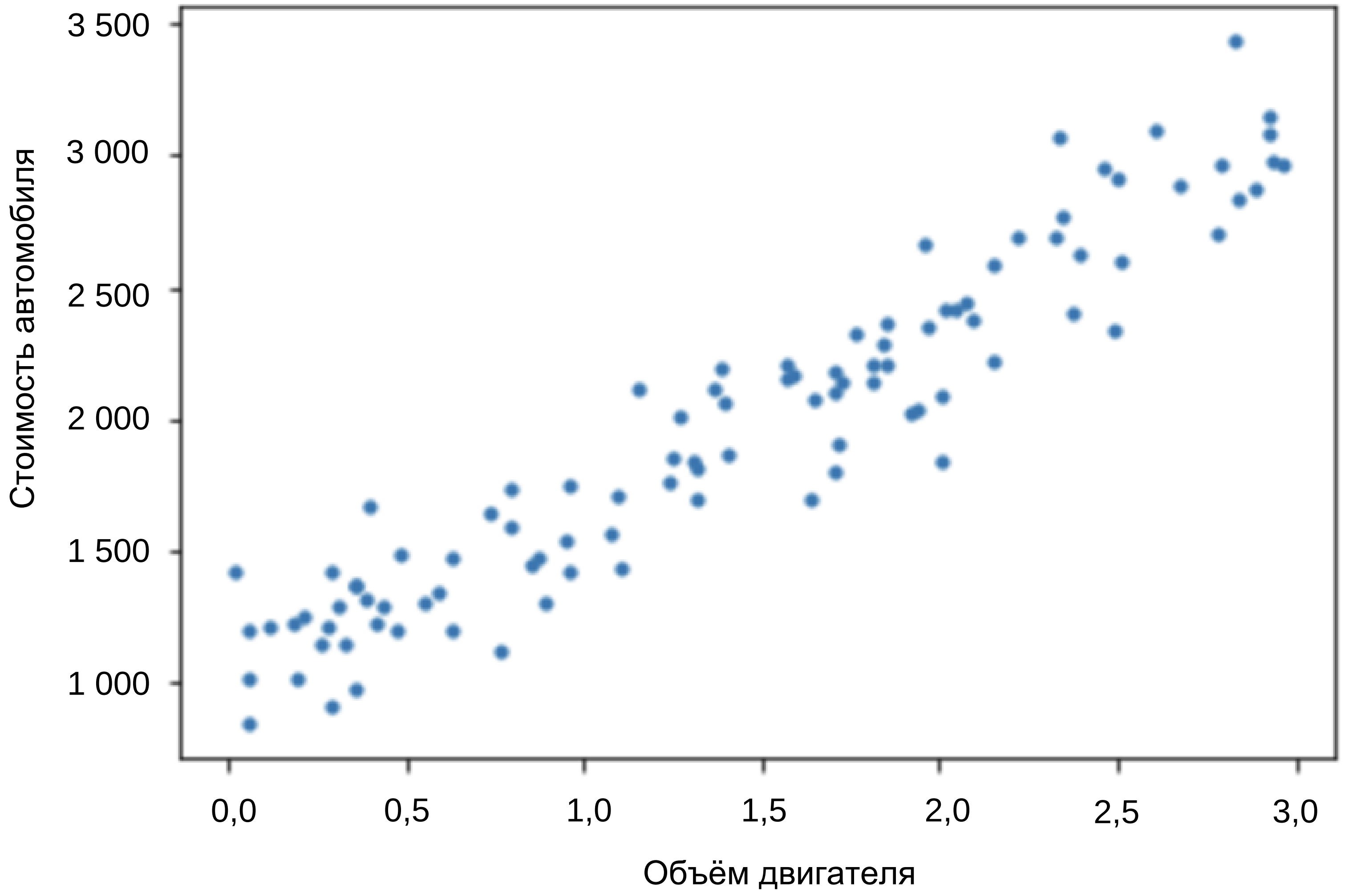
Уравнение регрессии при k = 1

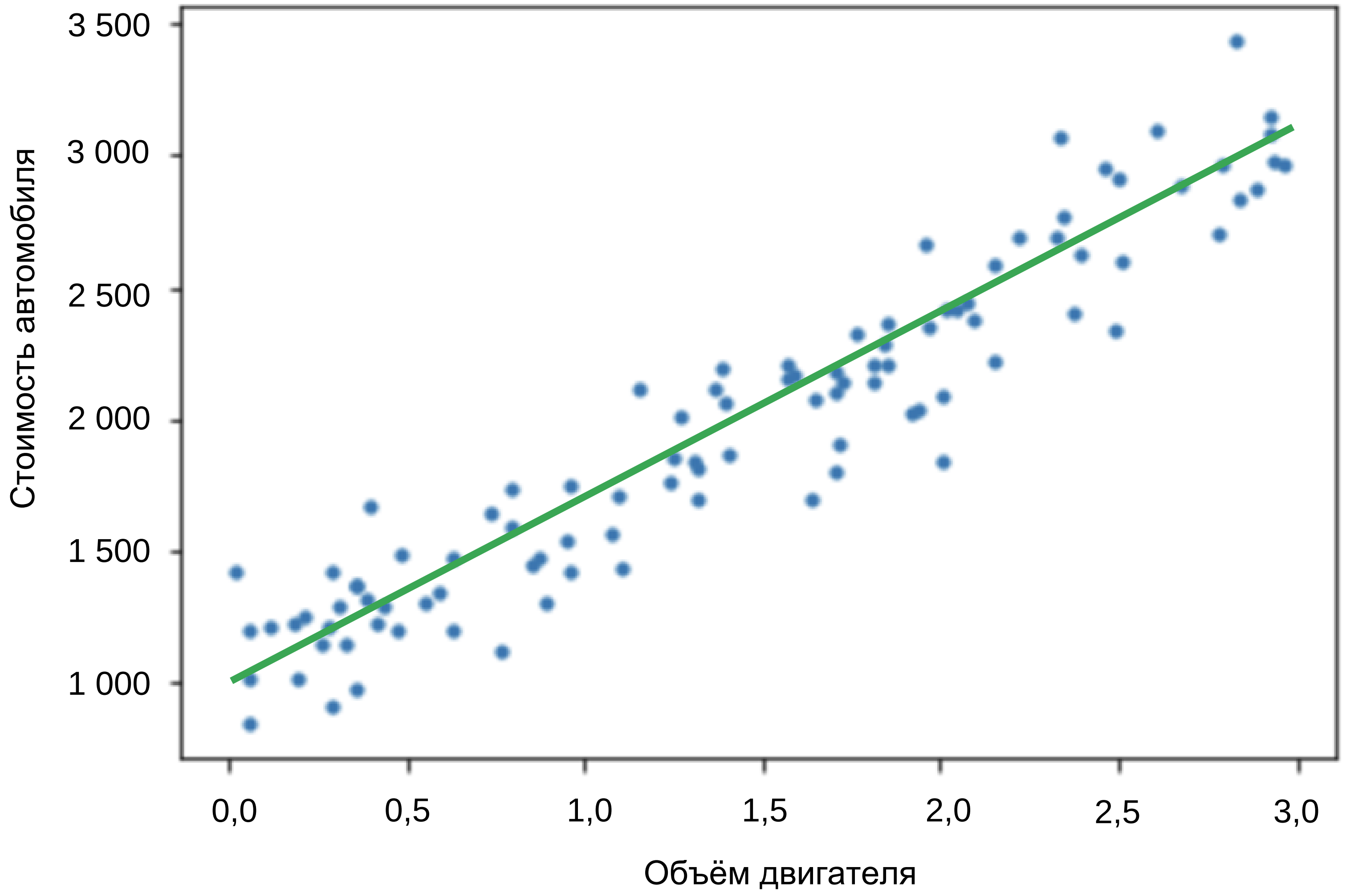
$$\hat{y} = w_0 + w_1 \times x_1$$

или

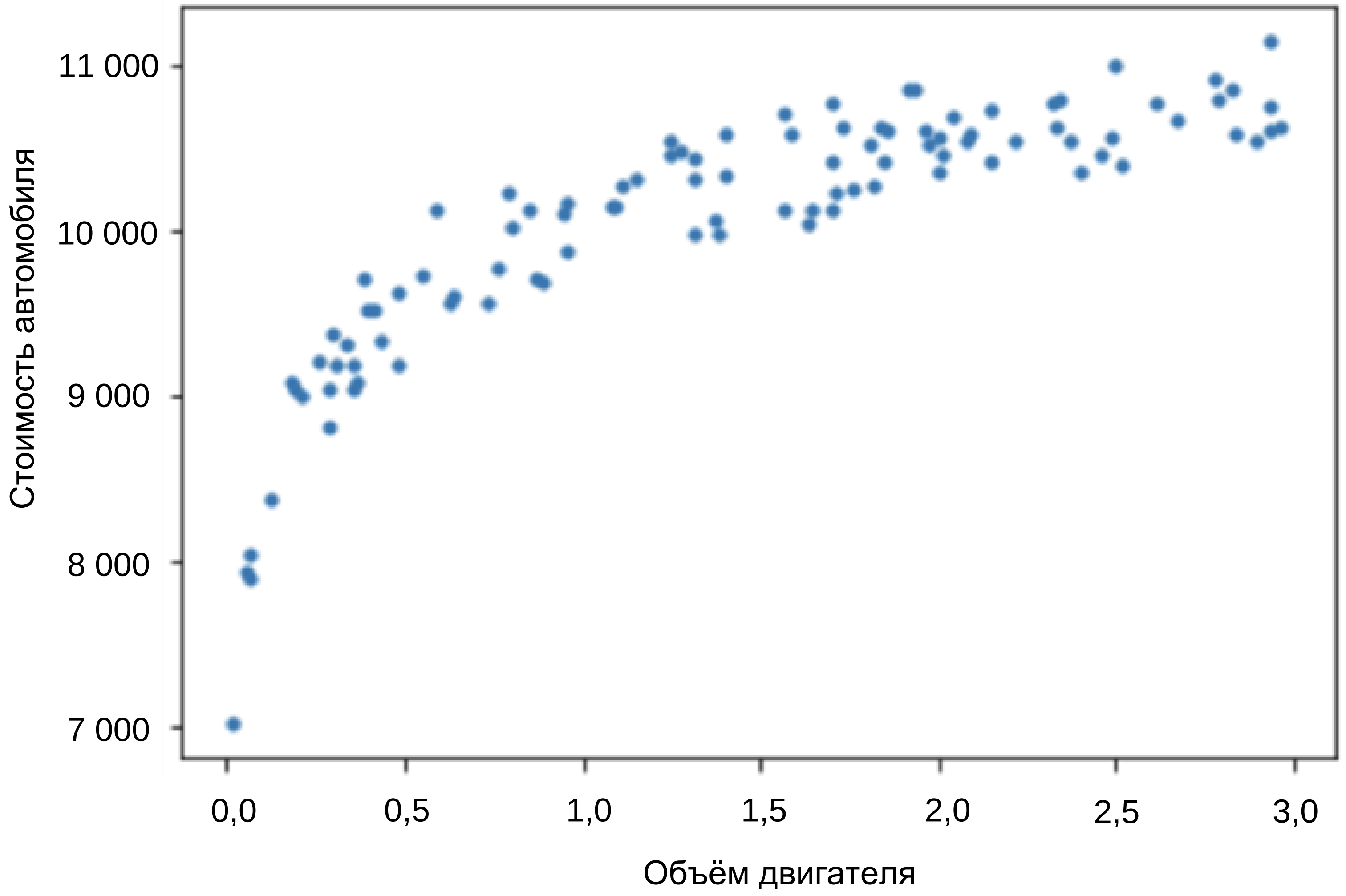
$$\hat{y} = a \times x + b$$

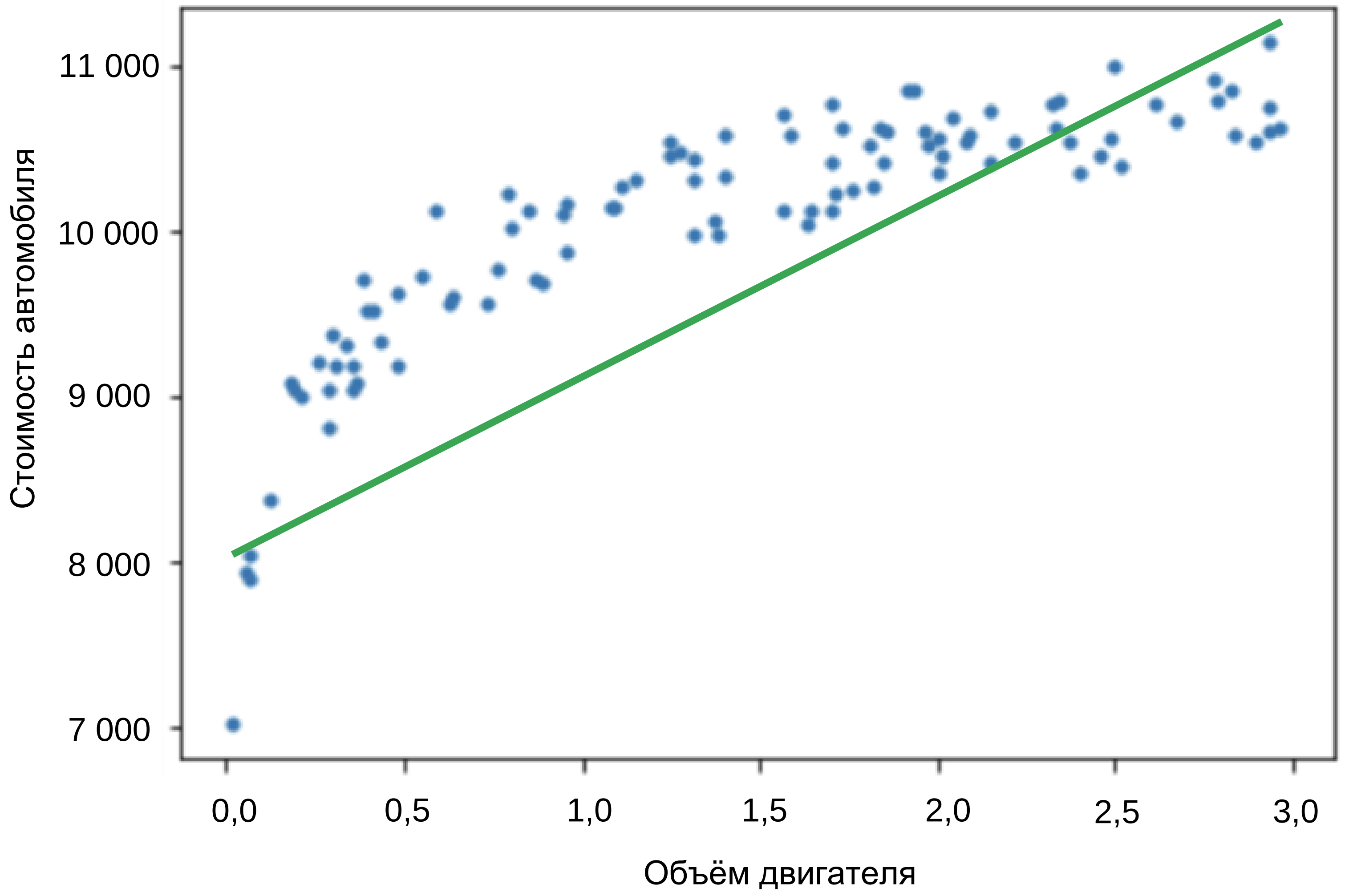
Это называется одномерная регрессия



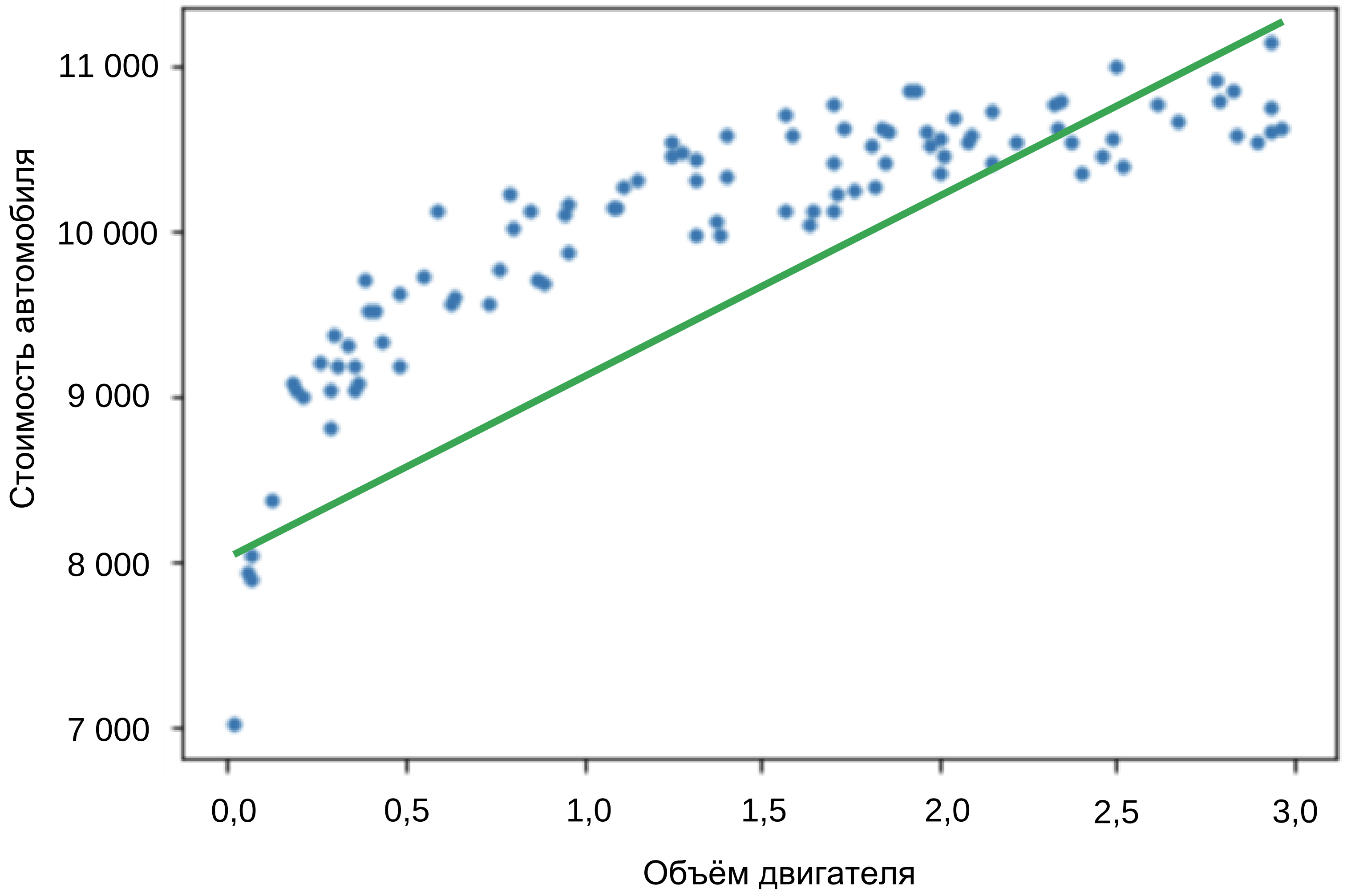


$$\hat{y} = w_0 + w_1 \times x_1$$

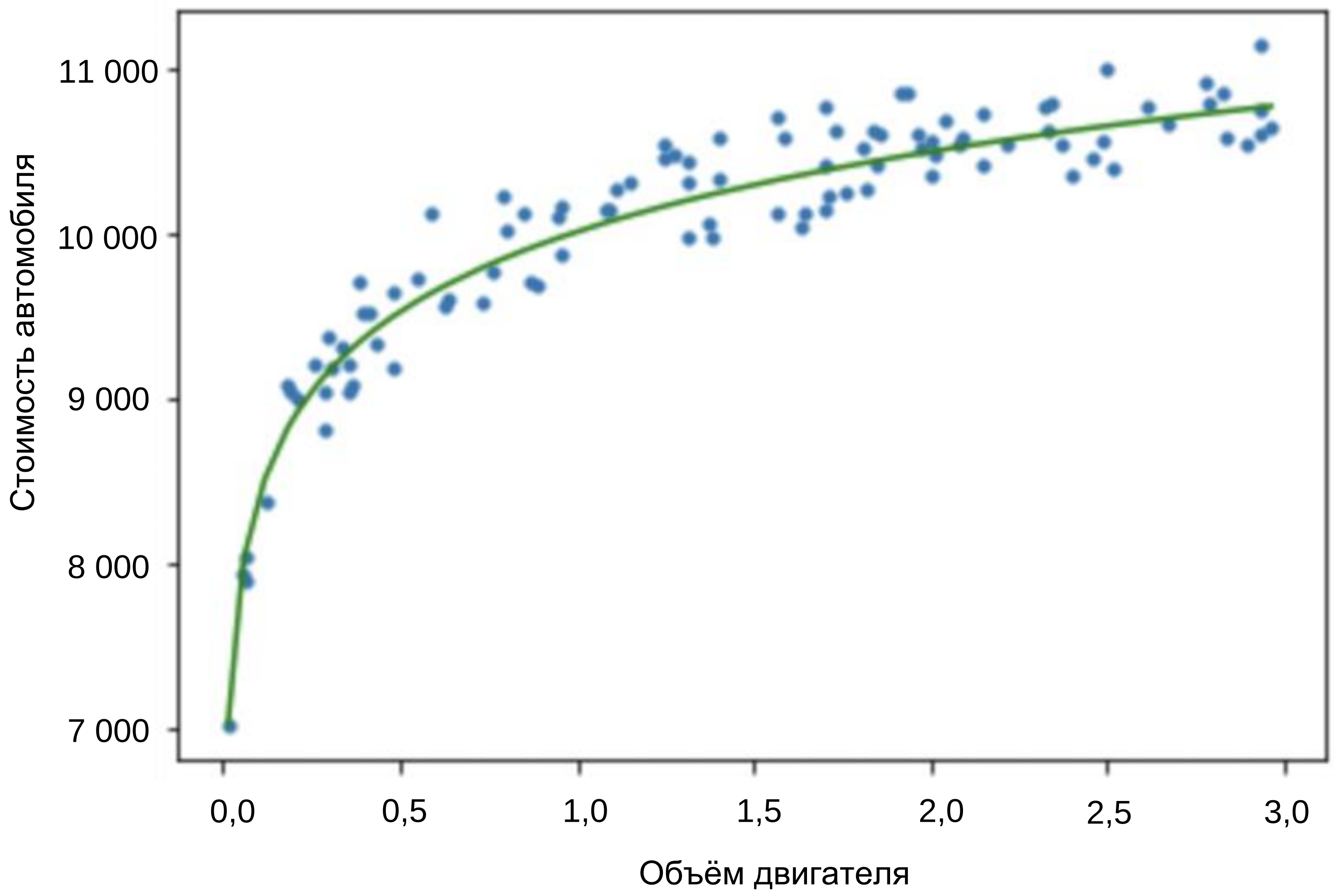




$$\hat{y} = w_0 + w_1 \times x_1$$

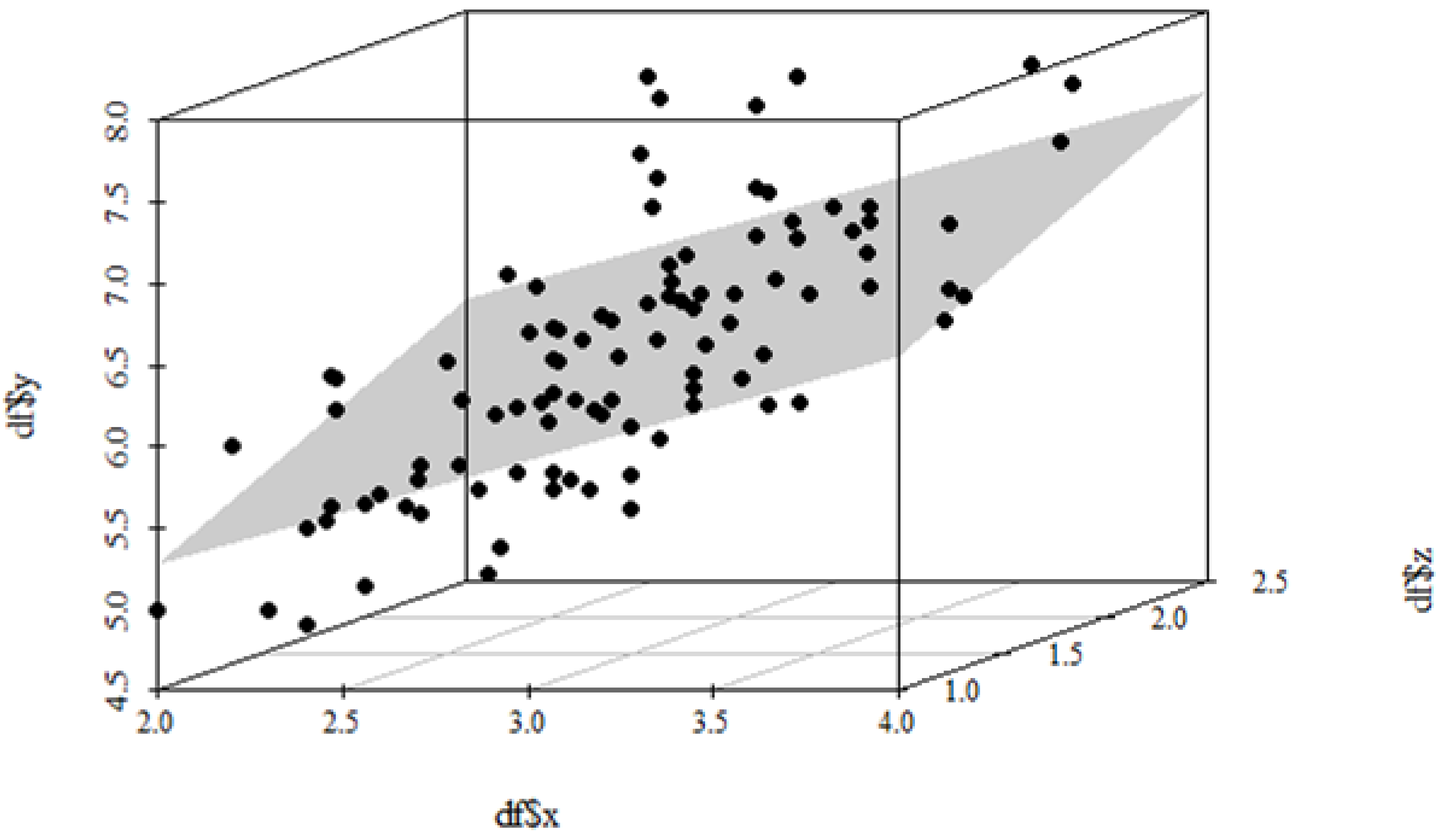


$$\hat{y} = w_0 + w_1 \times x_1$$



$$\hat{y} = w_0 + w_1 \times x_1 + w_2 \times \log(x_1)$$

Regression Plane



Уравнение регрессии

$$\hat{y} = f(w, x) = \boxed{w_0} + w_1 \times x_1 + \ldots + w_k \times x_k$$

$$\Downarrow$$

$$\hat{y} = f(w, x) = \boxed{w_0 \times 1} + w_1 \times x_1 + \ldots + w_k \times x_k$$

Для i-го объекта

Значения признаков наблюдений в качестве вектора.

Вектор будет иметь размерность 1xk, то есть 1 строка — один объект и k столбцов — признаков этого объекта:

$$(x_1, \ldots, x_k)$$

Коэффициенты регрессии

Тоже можно представить в виде вектора,
имеющего размерность kx1:

$$\vec{w} = \begin{pmatrix} w_1 \\ w_2 \\ \dots \\ w_k \end{pmatrix}$$

Уравнение регрессии

$$(x_{i1} \dots x_{iD}) \times \begin{pmatrix} w_1 \\ \cdots \\ w_D \end{pmatrix} + w_0 = (1\ x_{i1} \dots x_{iD}) \times \begin{pmatrix} w_0 \\ w_1 \\ \cdots \\ w_D \end{pmatrix}$$

Или

$$f(w, x_i) = x_i^T w$$

Уравнение регрессии

$$X = \begin{pmatrix} x_{00} & x_{01} & \dots & x_{0k} \\ x_{10} & x_{11} & \dots & x_{1k} \\ \dots & \dots & \dots & \dots \\ x_{n0} & x_{n1} & \dots & x_{nk} \end{pmatrix}$$

$$f(w, X) = Xw$$

$$\hat{y} = Xw$$

$$\vec{y} = \begin{pmatrix} y_0 \\ y_1 \\ \dots \\ y_n \end{pmatrix}$$