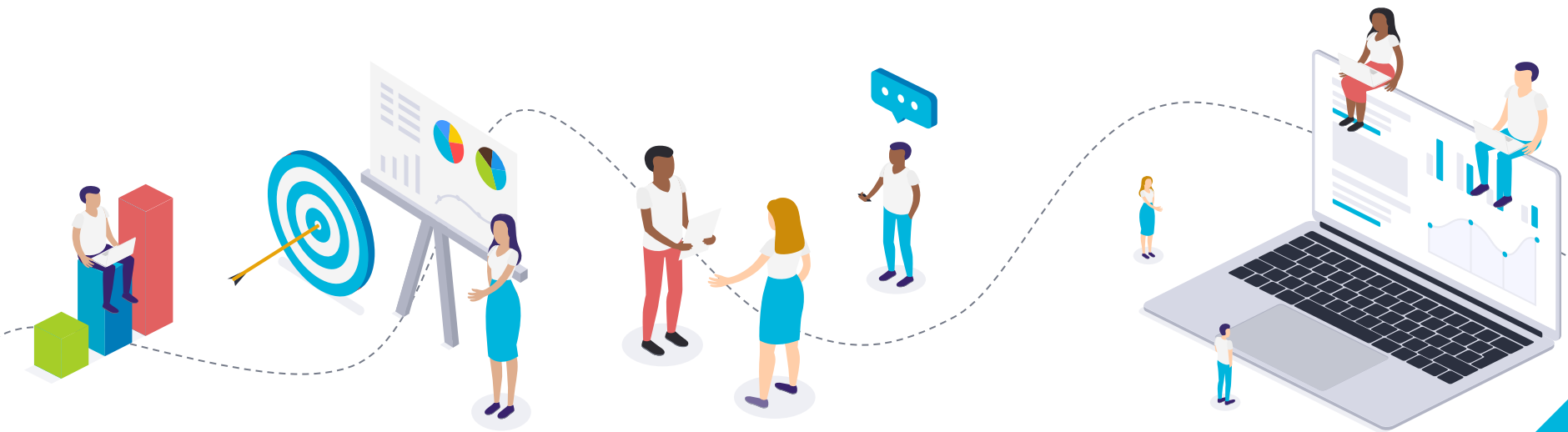
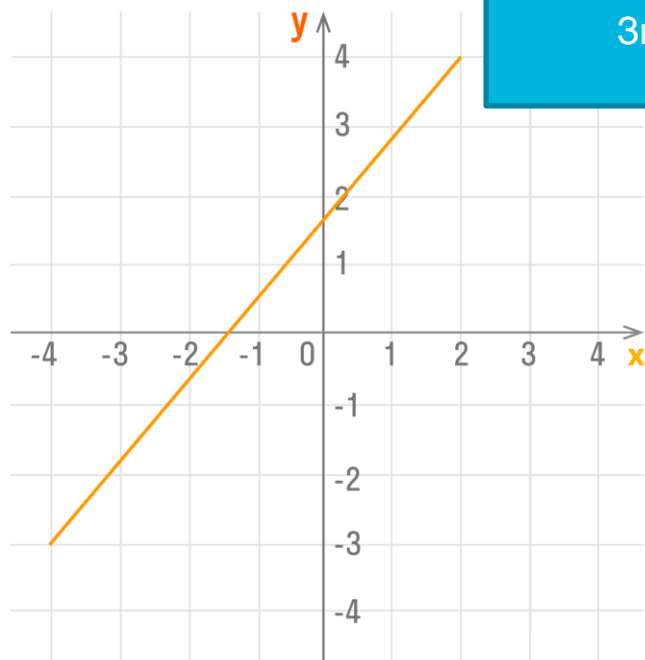


## Программа профессиональной переподготовки «Технологии искусственного интеллекта, визуализации и анализа данных»

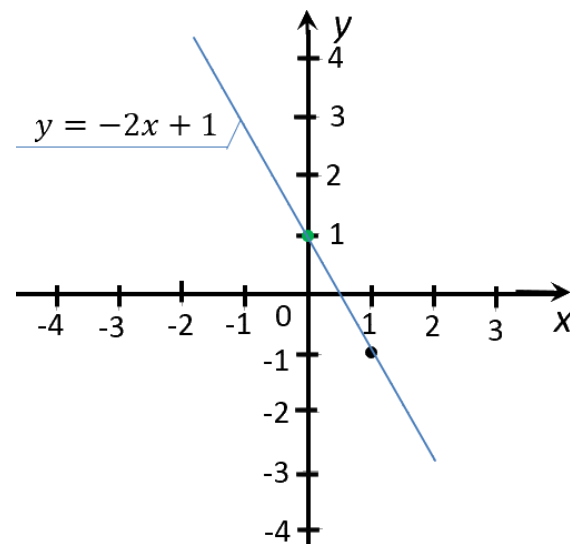


# Задача регрессии – предсказание значений непрерывной целевой переменной

График функции  $y = kx + b$



Знаком ли вам этот график?



# Линейная регрессия

Задача простой (одномерной) линейной регрессии состоит в том, чтобы смоделировать связь между единственным признаком (переменной  $x$ ) и целевой переменной (ответом)  $y$ .

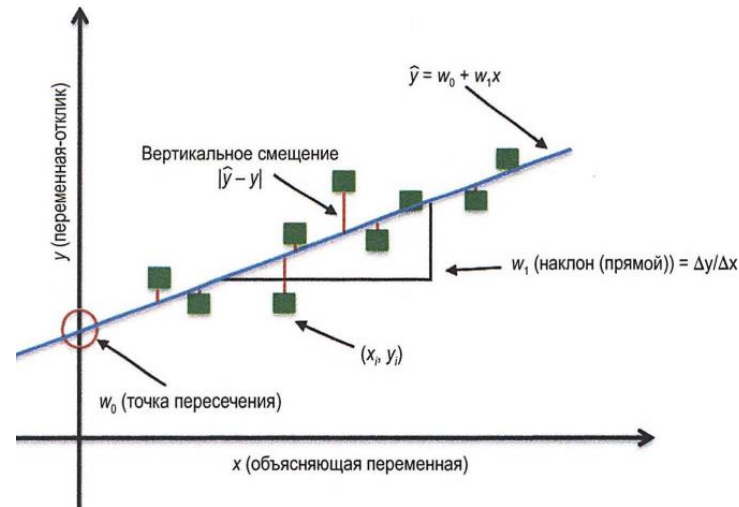
Уравнение линейной модели с одним признаком  $x$  определяется следующим образом:

$$y = \omega_0 + \omega_1 x.$$

Цель – определить  $\omega_0$  и  $\omega_1$  таким образом, чтобы описать связь между признаком  $x$  и целевой переменной  $y$ .

Линейная регрессия может пониматься как нахождение оптимально подогнанной прямой линии, проходящей через точки данных.

Оптимально подогнанная прямая линия также называется *линей регрессии*, а вертикальные прямые от линии регрессии до точек данных – это *ошибки предсказания*.



## Линейная регрессия

### Если признаков будет больше?

Случай, когда используется один признак, называется *простой линейной регрессией*, но, разумеется, мы также можем обобщить линейную регрессионную модель на два и более признака. Этот процесс называется *множественной линейной регрессией*.

$$\hat{y} = \alpha(x) = \omega_0 + \sum_{j=1}^d x_j \omega_j = \omega_0 + \langle x, \omega \rangle$$

$\omega_1, \dots, \omega_d$  - веса (или коэффициенты)

$\omega_0$  - сдвиг

Добавим признак  $x_0$  равный 1, это позволит еще упростить запись:

$$\hat{y} = \alpha(x) = \sum_{j=0}^d x_j \omega_j = \langle x, \omega \rangle$$

## Как найти $\omega$ ? Метод наименьших квадратов (МНК)

Суть МНК: сумма квадратов ошибок должна быть минимальной.

$$\sum_{i=1}^n (\alpha(x_i) - y_i)^2 \rightarrow \min$$

Продолжение в файле  
Linear Regression.ipynb

