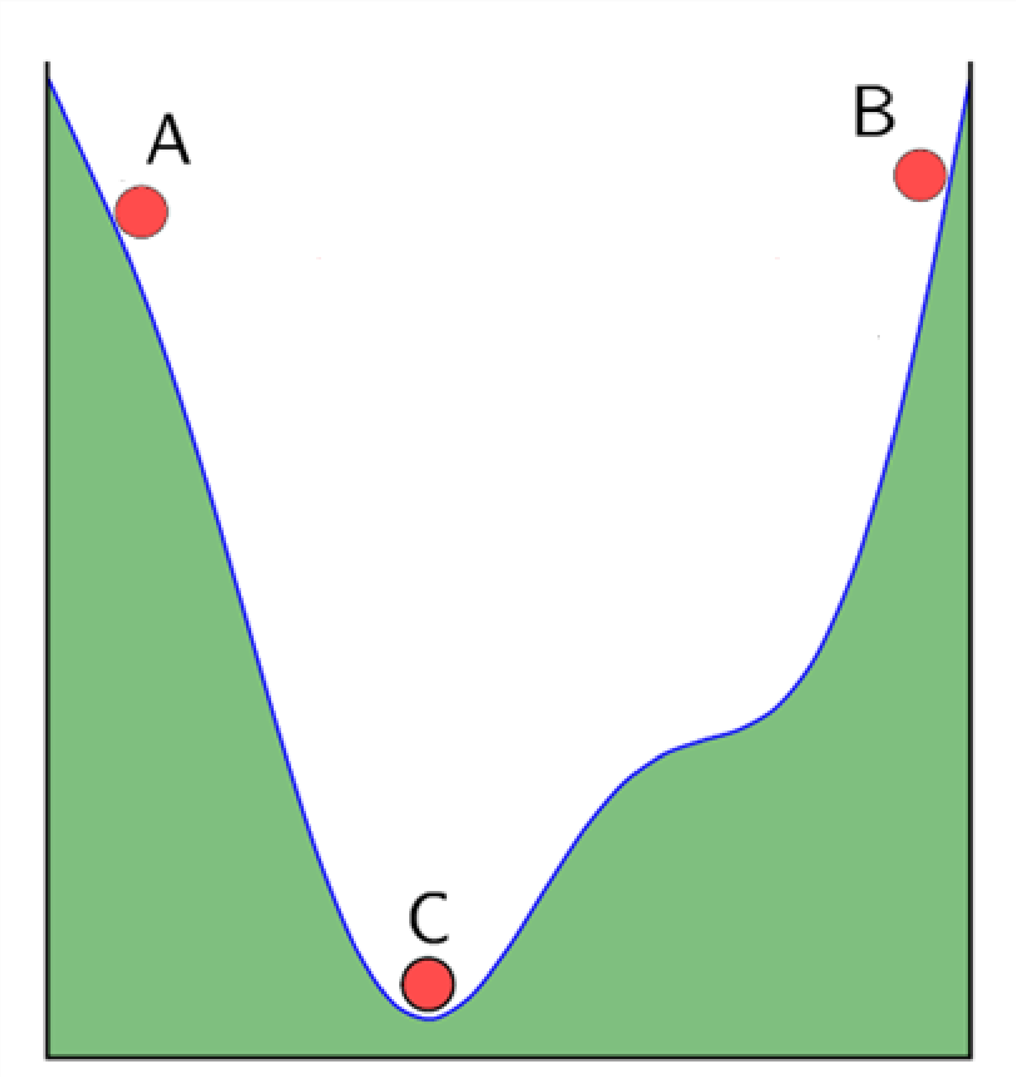
Градиентный спуск

Елена Кантонистова

Skillbox

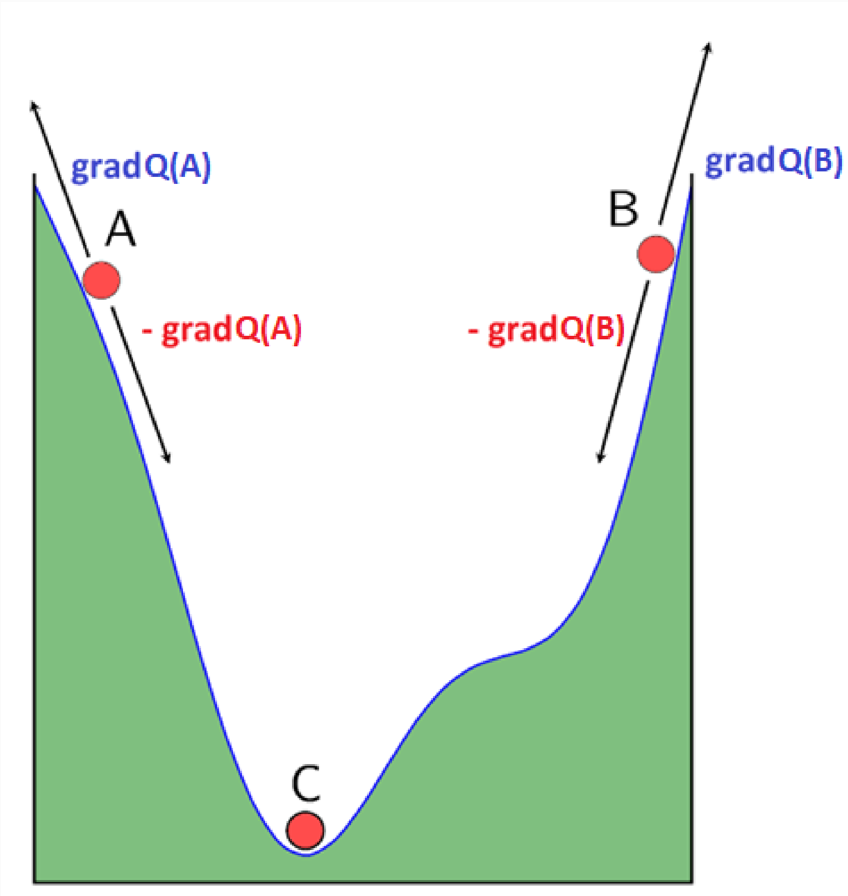
Как искать минимум функции?



Изображение: работа спикера

Градиент функции

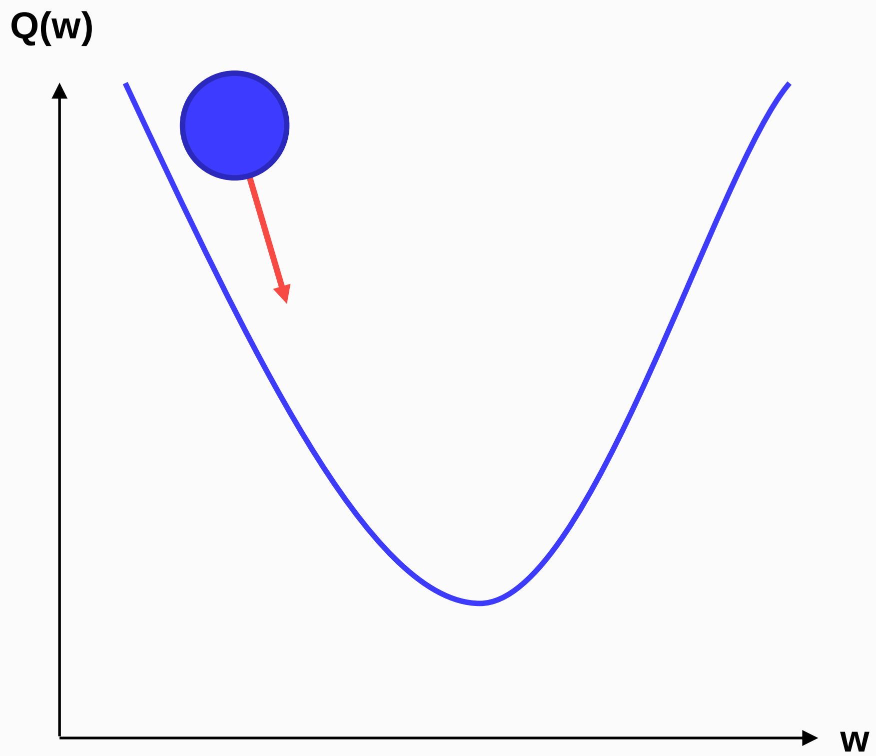
Градиент — вектор,
в направлении которого
функция быстрее всего
растёт.



Изображение: работа спикера

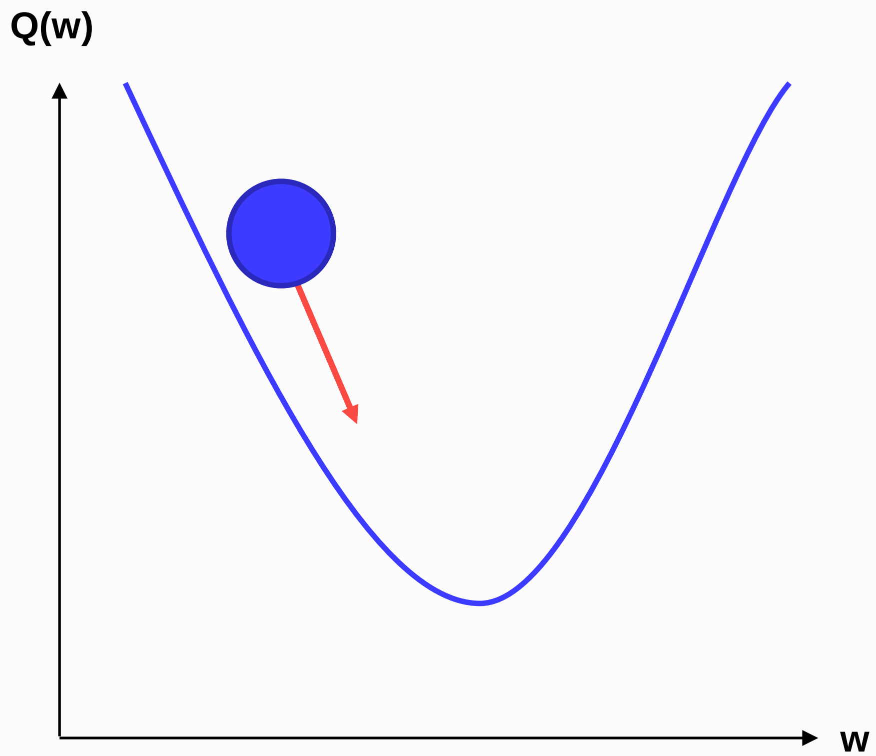
Градиент функции

Градиент — вектор,
в направлении которого
функция быстрее всего
растёт.



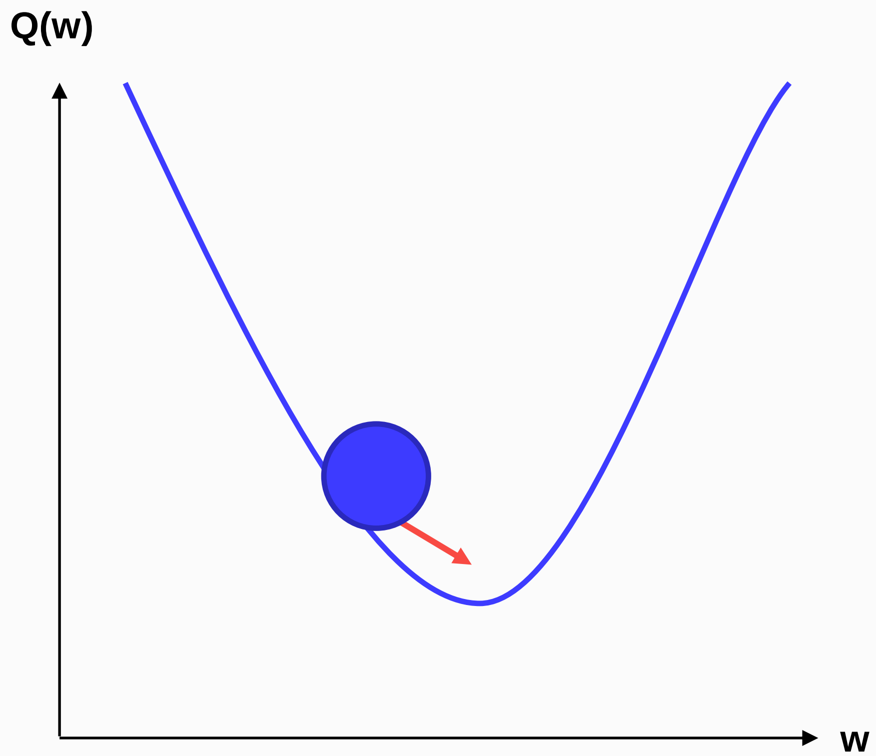
Градиент функции

Вычисляем градиент функции
в точке и сдвигаемся
в противоположном ему
направлении.



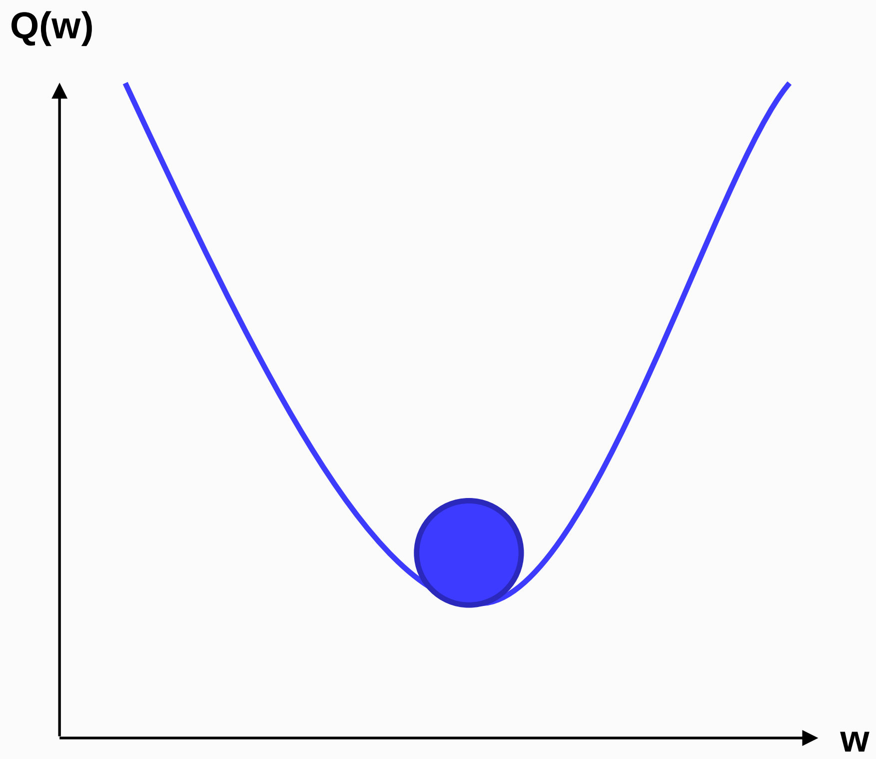
Пример

Вычисляем градиент функции в точке и сдвигаемся в противоположном ему направлении.



Пример

Вычисляем градиент функции в точке и сдвигаемся в противоположном ему направлении.

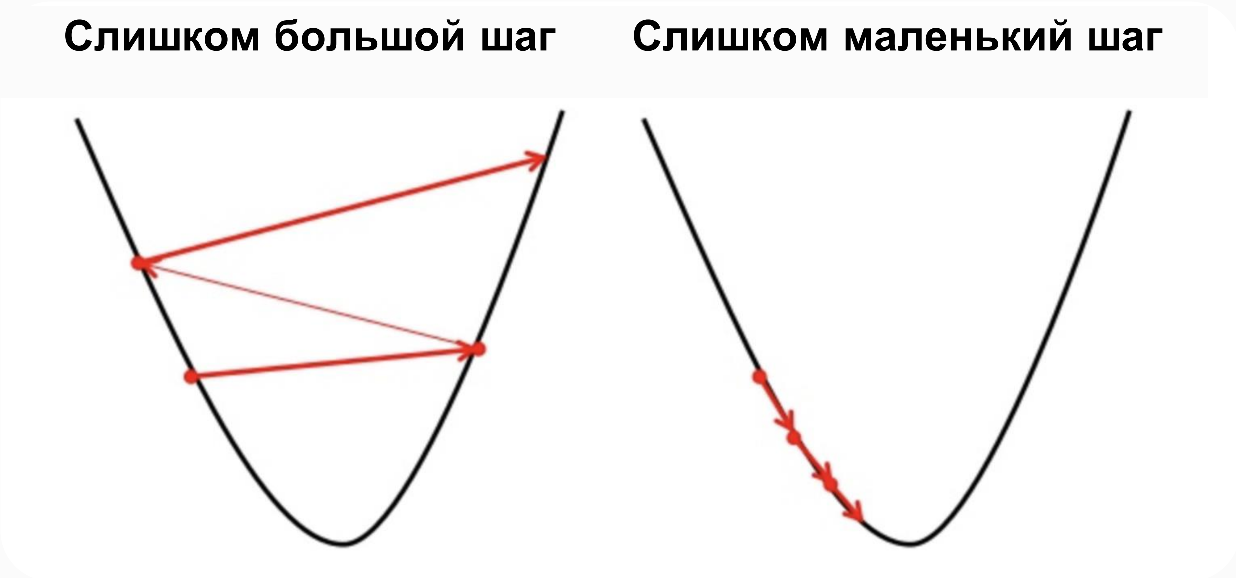


Формула градиентного спуска

Вычисляем градиент (в одномерном случае — производную) функции $Q(w)$ в точке и сдвигаемся в противоположном ему направлении:

$$w_{new} = w_{old} - Q'(w_{old})$$

Градиентный шаг



Добавляем в формулу гиперпараметр *a* — шаг градиентного спуска:

$$w_{new} = w_{old} - \alpha \cdot Q'(w_{old})$$

Градиент

- **Градиент** функции $Q(w)$ — это вектор

$$\nabla Q(w) = \{\frac{\partial Q}{\partial w_0}, \frac{\partial Q}{\partial w_1}, \ldots, \frac{\partial Q}{\partial w_d}\},$$

где $w = (w_0, w_1, \ldots, w_d)$

- Вектор — $\nabla Q(w)$, противоположный градиенту, называется **антиградиентом**

Формула градиентного спуска

В случае, если $Q(w)$ — функция многих переменных,
т. е. w — это не одно число, а вектор $w = (w_0, w_1, \ldots)$:

$$\boldsymbol{w}_{\boldsymbol{new}} = \boldsymbol{w}_{\boldsymbol{old}} - \alpha \cdot \nabla Q\left(\boldsymbol{w}_{\boldsymbol{old}}\right)$$

Недостатки градиентного спуска

$$\boldsymbol{w}_{new} = \boldsymbol{w}_{old} - \alpha \cdot \nabla Q\,(\boldsymbol{w}_{old})$$

На каждой итерации вычисляем градиент, т. е. производные по всем параметрам и для каждого объекта:

- большие временные затраты
- большие затраты по памяти

Стохастический градиентный спуск

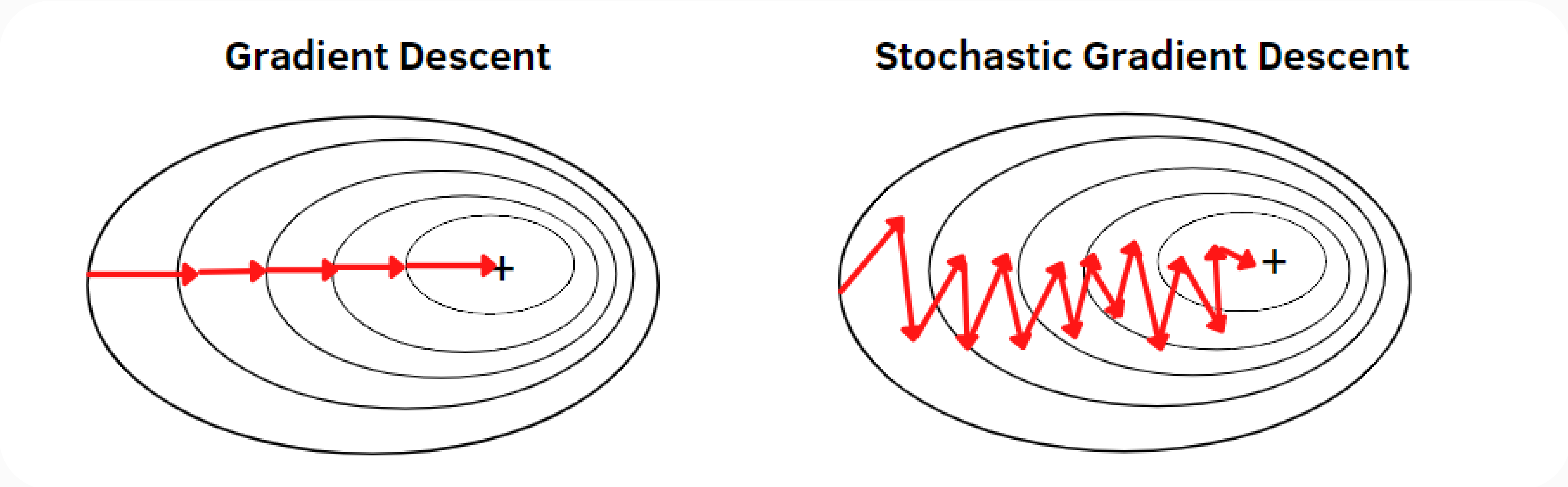
На каждой итерации:

- выбираем случайный объект обучающей выборки x_i
- вычисляем градиент только на этом объекте:

$$\boldsymbol{w}_{\boldsymbol{new}} = \boldsymbol{w}_{\boldsymbol{old}} - \alpha \cdot \nabla Q \left(\boldsymbol{w}_{\boldsymbol{old}}\right)$$

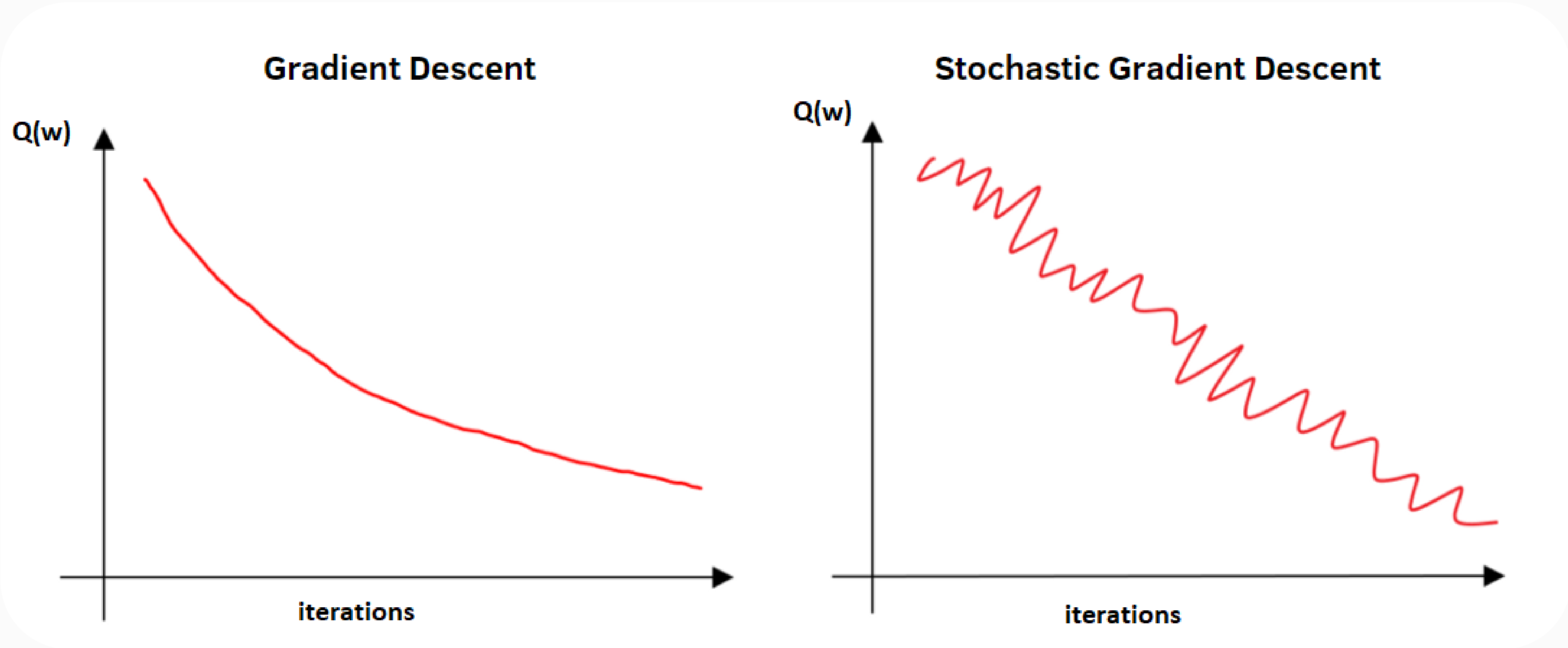
где $\nabla Q_i \left(\boldsymbol{w}_{\boldsymbol{old}}\right) = \nabla Q(\boldsymbol{w}_{\boldsymbol{old}}, x_i, y_i)$

GD vs SGD



Изображение: работа спикера

GD vs SGD



Изображение: работа спикера

Стохастический градиентный спуск

Утверждение. Метод стохастического градиентного спуска найдёт локальный минимум функции так же успешно, как и градиентный спуск, в случае, если минимум есть.

Стохастическому градиентному спуску может потребоваться больше итераций, однако он гораздо менее затратный по ресурсам.