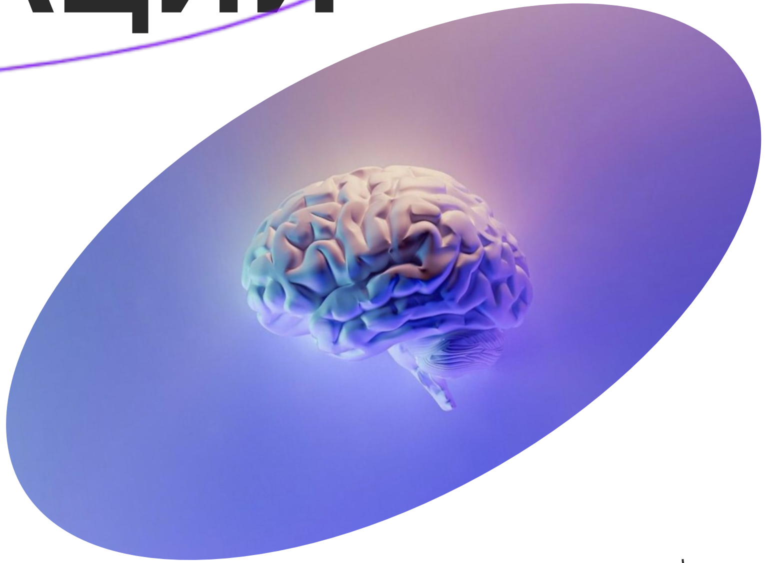
ЗАДАЧА КЛАССИФИКАЦИИ



БИНАРНАЯ КЛАССИФИКАЦИЯ

Датасет $X \in \mathbb{R}^{m \times n}$

Классы $Y \in \{0, 1\}$.

$$X = \begin{pmatrix} x_{11} & x_{12} & \dots & x_{1n} \\ x_{21} & x_{22} & \dots & x_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ x_{m1} & x_{m2} & \dots & x_{mn} \end{pmatrix} \quad Y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_m \end{pmatrix}$$

БИНАРНАЯ КЛАССИФИКАЦИЯ

Пример

знаем данные пассажиров титаника, надо предсказать выжил он или нет.

Возраст	Класс билета	Пассажир выжил
46	A	1
20	B	0
65	B	0
37	A	1
12	B	1
39	B	0
57	A	1

ВЕРОЯТНОСТНАЯ МОДЕЛЬ

X - случайная величина, вектор признаков.

Y - случайная величина, целевая переменная.

Пример: клик на рекламу.

$X =$ (кол-во кликов раньше, время активности, уровень доходов)

$Y = 1$, если кликнет, 0 иначе

$P(Y = 1|X)$ - распределение вероятностей.

ВЕРОЯТНОСТНАЯ МОДЕЛЬ

Хотим уметь моделировать $P(Y=1|X)$.

Так как у нас задача бинарной классификации можно моделировать только $P(Y=1|X)=f(x)$.

$f(x)$ – модель.

Вероятность принадлежности к 0 классу будет $P(Y=0|X)=1-P(Y=1|X)$.

ПРАВДОПОДОБИЕ

Правдоподобие – произведение вероятностей появления

каждого элемента из выборки.

Это будет вероятность получения нашей выборки.

$$\prod_{i=1}^{m} P(Y = y_i|x_i) \rightarrow max$$

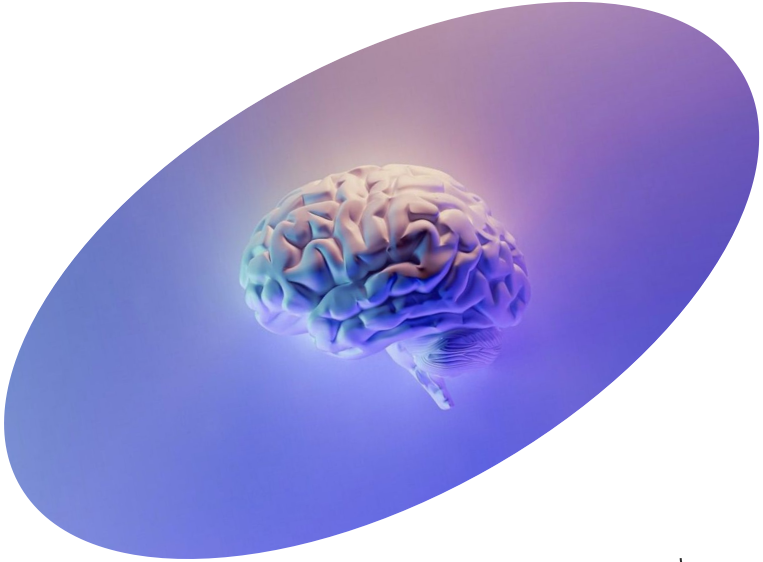
ПРАВДОПОДОБИЕ

$$\ln\left(\prod_{i=1}^{m} P(Y = y_i|x_i)\right) \to max$$

$$\sum_{i=1}^{m} \ln(P(Y = y_i|x_i)) \to max$$

$$L(w) = -\sum_{i=1}^{m} \ln(P(Y = y_i|x_i)) \to min$$

ЛОГИСТИЧЕСКАЯ РЕГРЕССИЯ

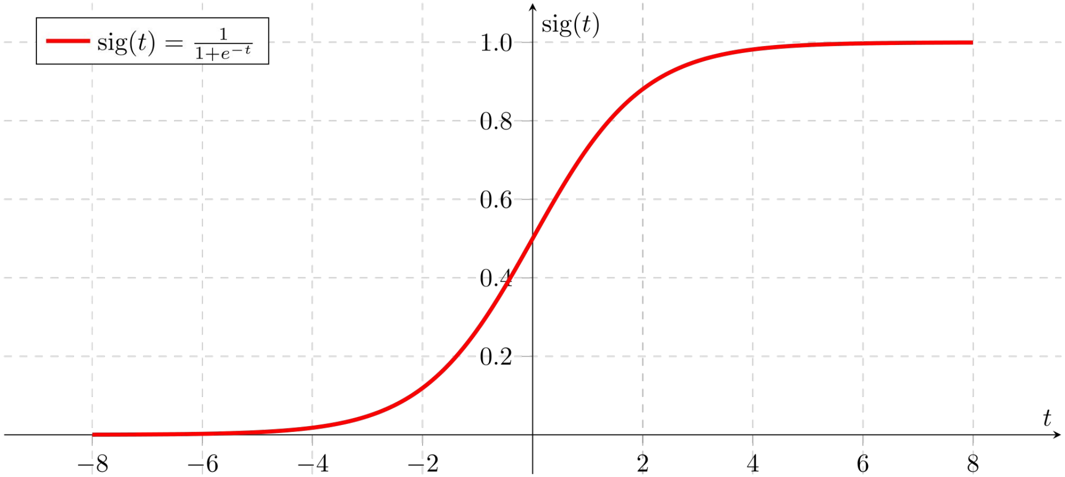


МОДЕЛЬ

$$f(x) = \sum_{i=1}^{n} w_i x_i = \mathbf{xw}$$

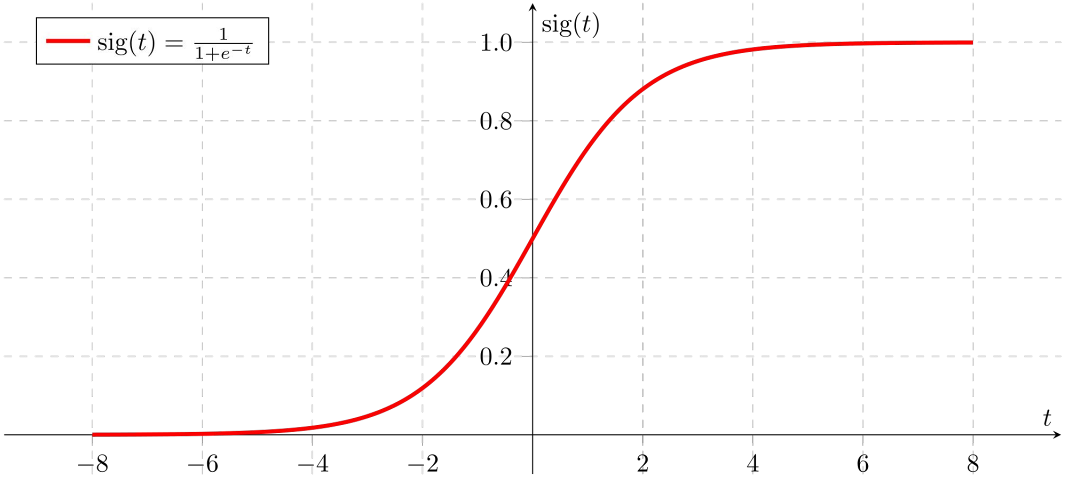
МОДЕЛЬ

$$\sigma(t) = \frac{1}{1 + e^{-t}}$$

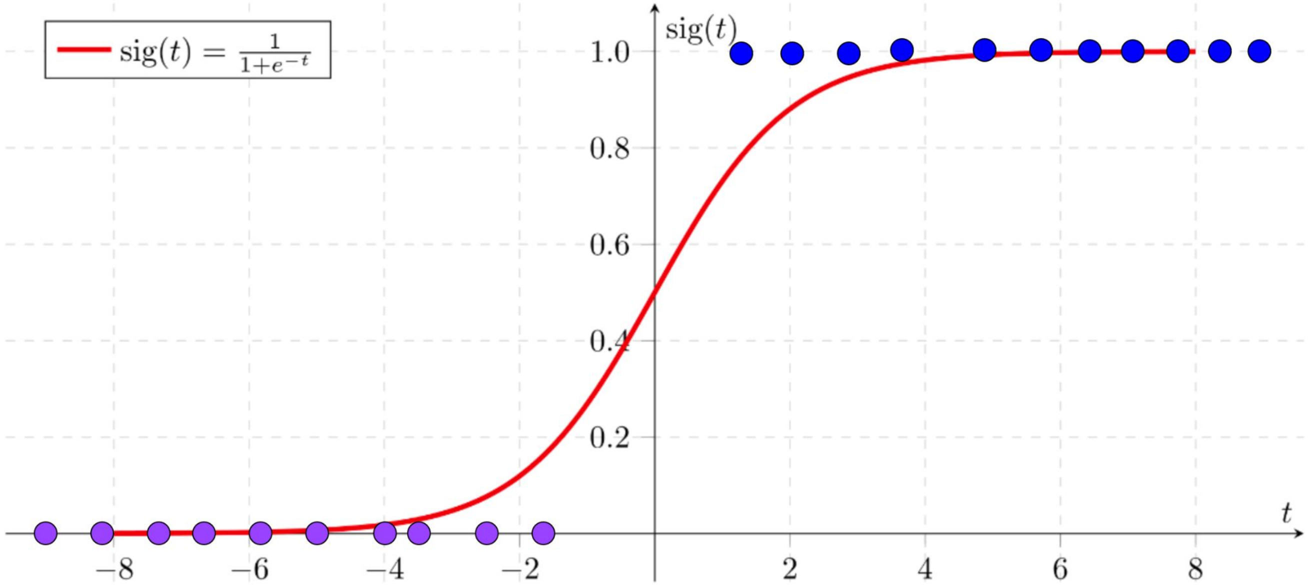


МОДЕЛЬ

$$f(x) = \sigma\left(\sum_{i=1}^{n} wx\right)$$



МОДЕЛЬ



БИНАРНАЯ КЛАССИФИКАЦИЯ LogReg

Датасет $X \in \mathbb{R}^{m \times n}$.

Классы $Y \in \{0, 1\}$.

$$f(x) = \sigma\left(\sum_{i=1}^{n} wx\right)$$

$$L(w) = -\sum_{i=1}^{m} \ln(P(Y = y_i | x_i)) \to min$$

БИНАРНАЯ КЛАССИФИКАЦИЯ LogReg

$$f(x) = \sigma\left(\sum_{i=1}^{n} wx\right)$$

$$L(w) = -\sum_{i=1}^{m} \ln(P(Y = y_i | x_i)) \to min$$

$$\ln P(Y = y_i | x_i) = y_i \ln(f(x_i)) + (1 - y_i) \ln(1 - f(x))$$

LogLoss

$$L(w) = -\sum_{i=1}^{m} y_i \ln(f(x_i)) + (1 - y_i) \ln(1 - f(x_i)) \to min$$

МНОГОКЛАССОВАЯ КЛАССИФИКАЦИЯ

$$logits(x) = \begin{pmatrix} w^1 x \\ w^2 x \\ \dots \\ w^m x \end{pmatrix} = \begin{pmatrix} \sum\limits_{i=0}^{n} w_i^1 x_i \\ \sum\limits_{i=0}^{n} w_i^2 x_i \\ \dots \\ \sum\limits_{i=0}^{n} w_i^m x_i \end{pmatrix}$$

SOFTMAX

$$softmax(\alpha) = \left(\frac{e^{\alpha_1}}{\sum e^{\alpha_i}}, \frac{e^{\alpha_2}}{\sum e^{\alpha_i}}, \ldots, \frac{e^{\alpha_m}}{\sum e^{\alpha_i}} \right)$$

$$\alpha_i = w^i x \ (\alpha - \text{вектор logits})$$

ПРИМЕР SOFTMAX

$$logits(x) = \begin{pmatrix} 1,3 \\ 5,1 \\ 2,2 \\ 0,7 \\ 1,1 \end{pmatrix} \longrightarrow \mathrm{softmax} \longrightarrow \begin{pmatrix} 0,02 \\ 0,90 \\ 0,05 \\ 0,01 \\ 0,02 \end{pmatrix}$$

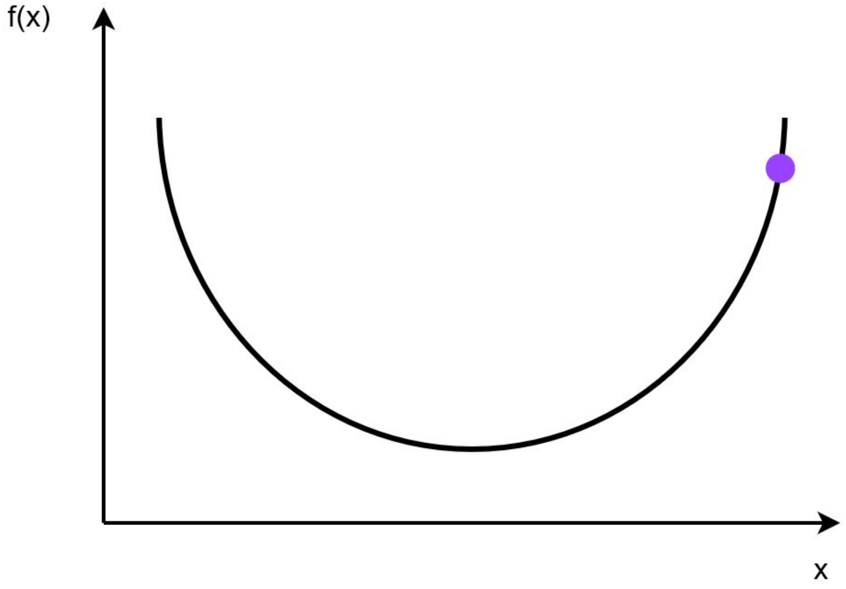
МНОГОКЛАССОВАЯ КЛАССИФИКАЦИЯ

$$f(x) = softmax(logits(x))$$

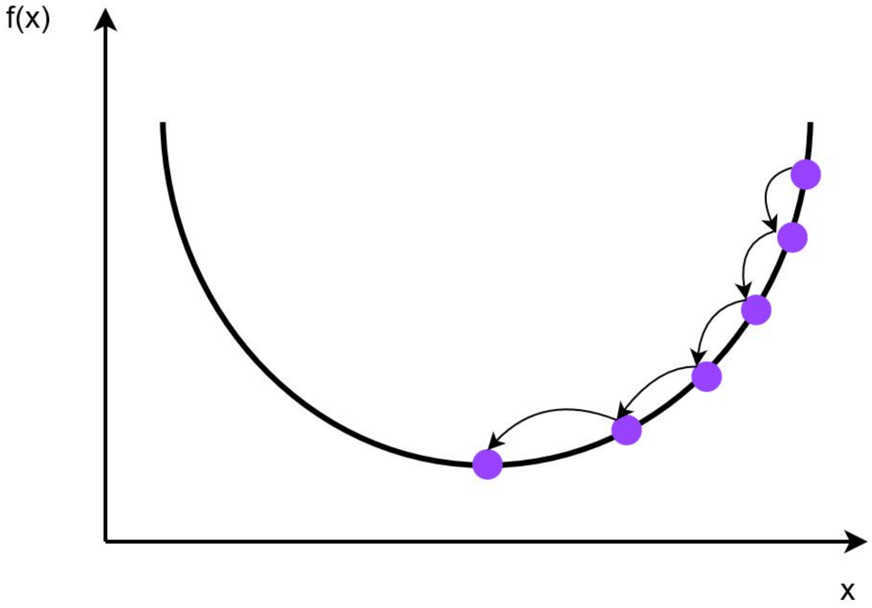
$$P(y = k|x) = \frac{e^{\sum w_i^k x_i}}{\sum e^{\sum w_i^k x_i}}$$

КАК МИНИМИЗИРОВАТЬ ФУНКЦИЮ ПОТЕРЬ?

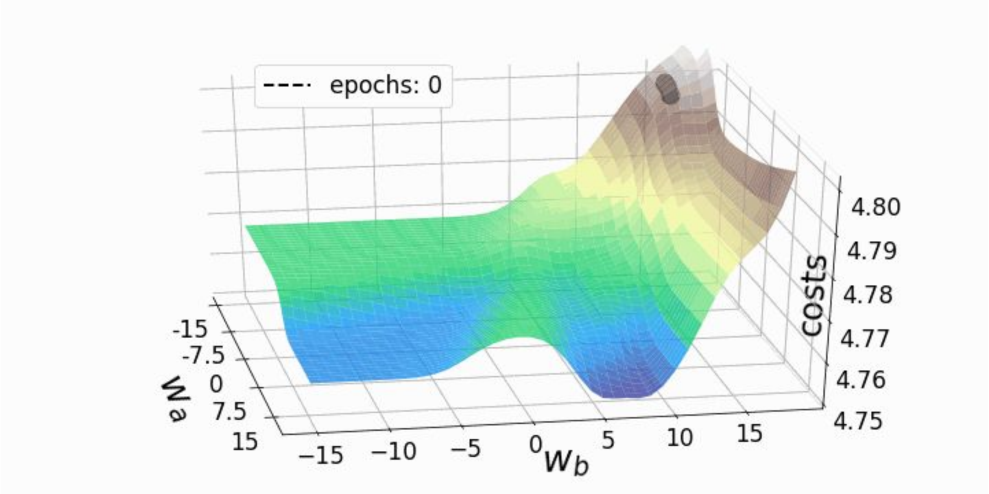
ПРИМЕР

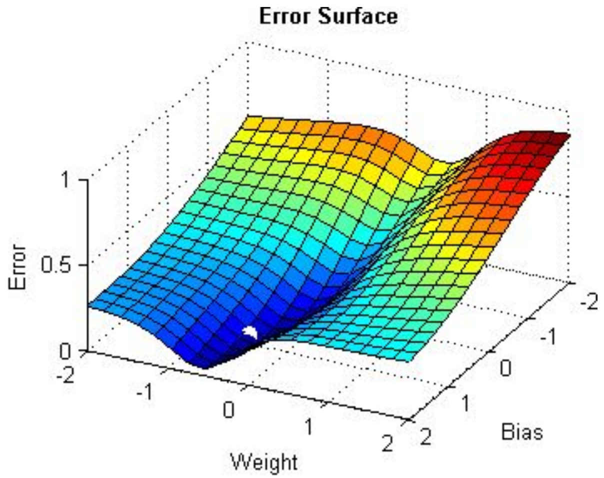


ПРИМЕР

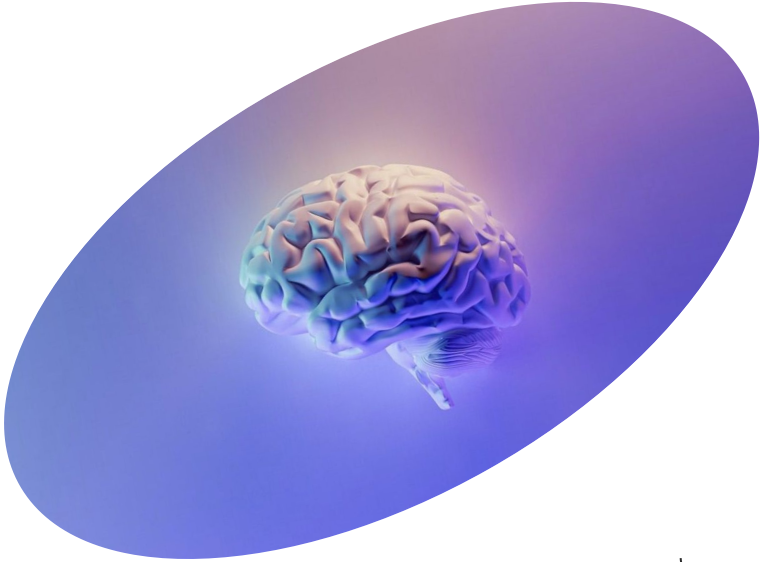


ПРИМЕР


epochs: 0
costs
4.80
4.79
4.78
4.77
4.76
4.75
w_a
-15
-7.5
0
7.5
15
w_b
−15
−10
−5
0
5
10
15


Error Surface
Error
1
0.5
0
Weight
-2
-1
0
1
2
Bias
-2
-1
0
1
2

ГРАДИЕНТНЫЙ СПУСК



ГРАДИЕНТНЫЙ СПУСК

Градиентный спуск – метод нахождения минимума функции (аргументов в которых функция принимает минимальное значение).

Надо найти веса, которые минимизируют функцию.

ГРАДИЕНТНЫЙ СПУСК

Градиентный спуск – метод нахождения минимума функции

Надо найти веса, которые минимизируют функцию.

$$f(x) \to min$$

learning rate (lr) – размер шага градиентного спуска

ГРАДИЕНТНЫЙ СПУСК

Для каждой переменной:

1) Выбираем начальную точку произвольно
2) Вычисляем значение градиента в точке
3) Домножаем на **learning rate**
4) Вычитаем из старого значения переменной

ГРАДИЕНТНЫЙ СПУСК

$$x_1 = x_1 - lr\frac{\partial f}{\partial x_1} \quad x_2 = x_2 - lr\frac{\partial f}{\partial x_2} \quad \ldots \quad x_n = x_n - lr\frac{\partial f}{\partial x_n}$$

$$x = x - lr\nabla f$$

ПРИМЕР

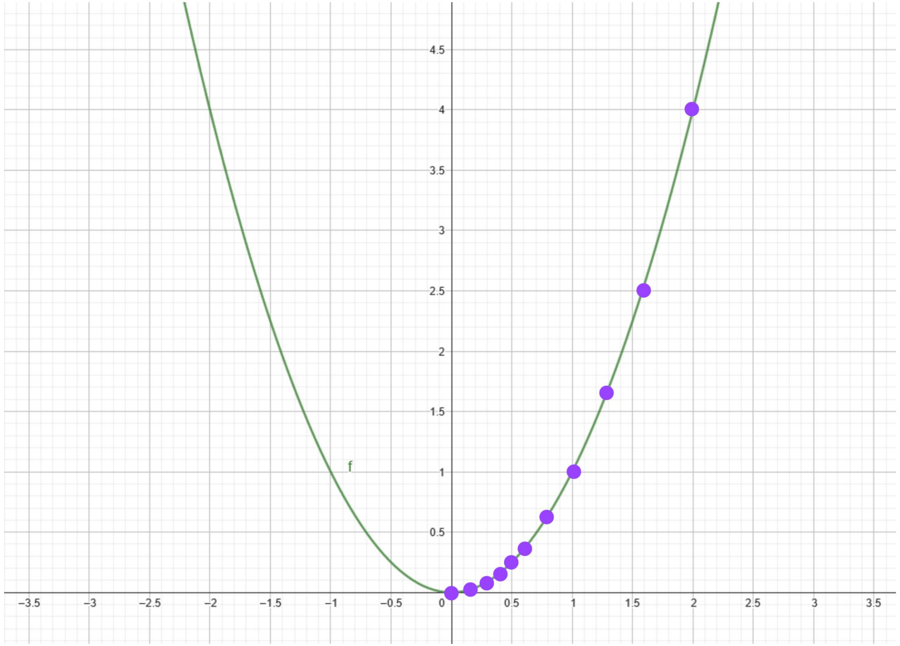
$$\text{Обновления: } x = x - 0,1 \cdot 2x$$

$$f(x) = x^2$$

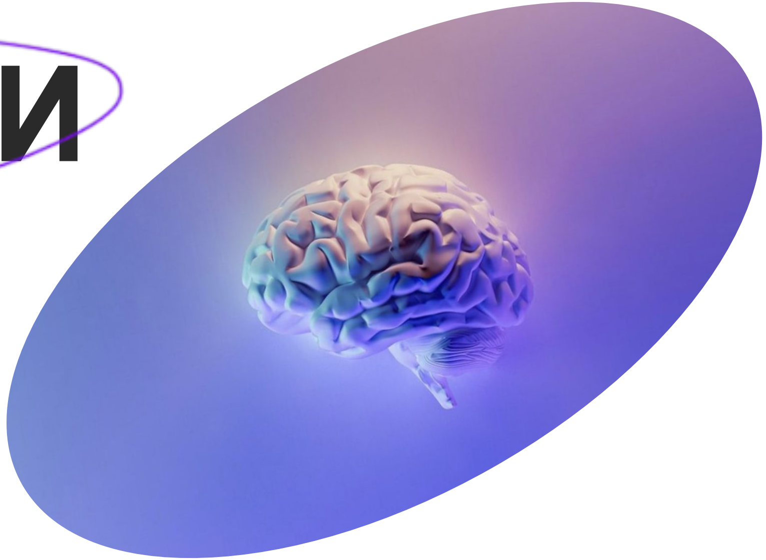
$$\nabla f(x) = 2x$$

$$lr = 0,1$$

$$x = 2$$



ГРАДИЕНТНЫЙ СПУСК ДЛЯ РЕГРЕССИИ



ФУНКЦИЯ ПОТЕРЬ ЛИНЕЙНОЙ РЕГРЕССИИ

$$L(w) = \sum (x_i^T w - y_i)^2$$

ГРАДИЕНТНЫЙ СПУСК ДЛЯ ЛИНЕЙНОЙ РЕГРЕССИИ

$$\frac{\partial L}{\partial w_1} = \sum \frac{(x_i^T w - y_i)^2}{\partial w_1}$$

$$= \sum 2(x_i^T w - y_i)^2 \frac{\partial x_i^T w}{\partial w_1}$$

$$= \sum 2(x_i^T w - y_i)^2 x_{i1}$$

ГРАДИЕНТНЫЙ СПУСК ДЛЯ ЛИНЕЙНОЙ РЕГРЕССИИ

Минимизируем функцию: $L(w) = \sum (x_i^T w - y_i)^2$

1. Произвольно выбираем начальные веса

2. Обновляем веса: $w_i = w_i - lr\frac{\partial L}{\partial w_i}$

ГРАДИЕНТНЫЙ СПУСК ДЛЯ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

$$L(w) = -\sum_{i=1}^{m} y_i \ln(f(x_i)) + (1 - y_i) \ln(1 - f(x_i)) \to min$$

ГРАДИЕНТНЫЙ СПУСК ДЛЯ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

$$L(w) = -\sum_{i=1}^{m} y_i \ln(f(x_i)) + (1 - y_i) \ln(1 - f(x_i)) \to min$$

$$\frac{\partial L}{\partial w_1} = -\sum \frac{y_i}{f(x_i)} \frac{\partial f(x_i)}{\partial w_1} - \frac{1 - y_i}{1 - f(x_i)} \frac{\partial f(x_i)}{\partial w_1}$$

ГРАДИЕНТНЫЙ СПУСК ДЛЯ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

$$f(x) = \sigma\left(\sum_{i=1}^{n} wx\right)$$

$$\frac{\partial f(x_i)}{\partial w_1} = f(x_i)(1 - f(x_i))x_{i1}$$

ГРАДИЕНТНЫЙ СПУСК ДЛЯ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ

$$\frac{\partial L}{\partial w_1} = -\sum x_{i1}\Big(y_i(1 - f(x_i)) - (1 - y_i)f(x_i)\Big)$$