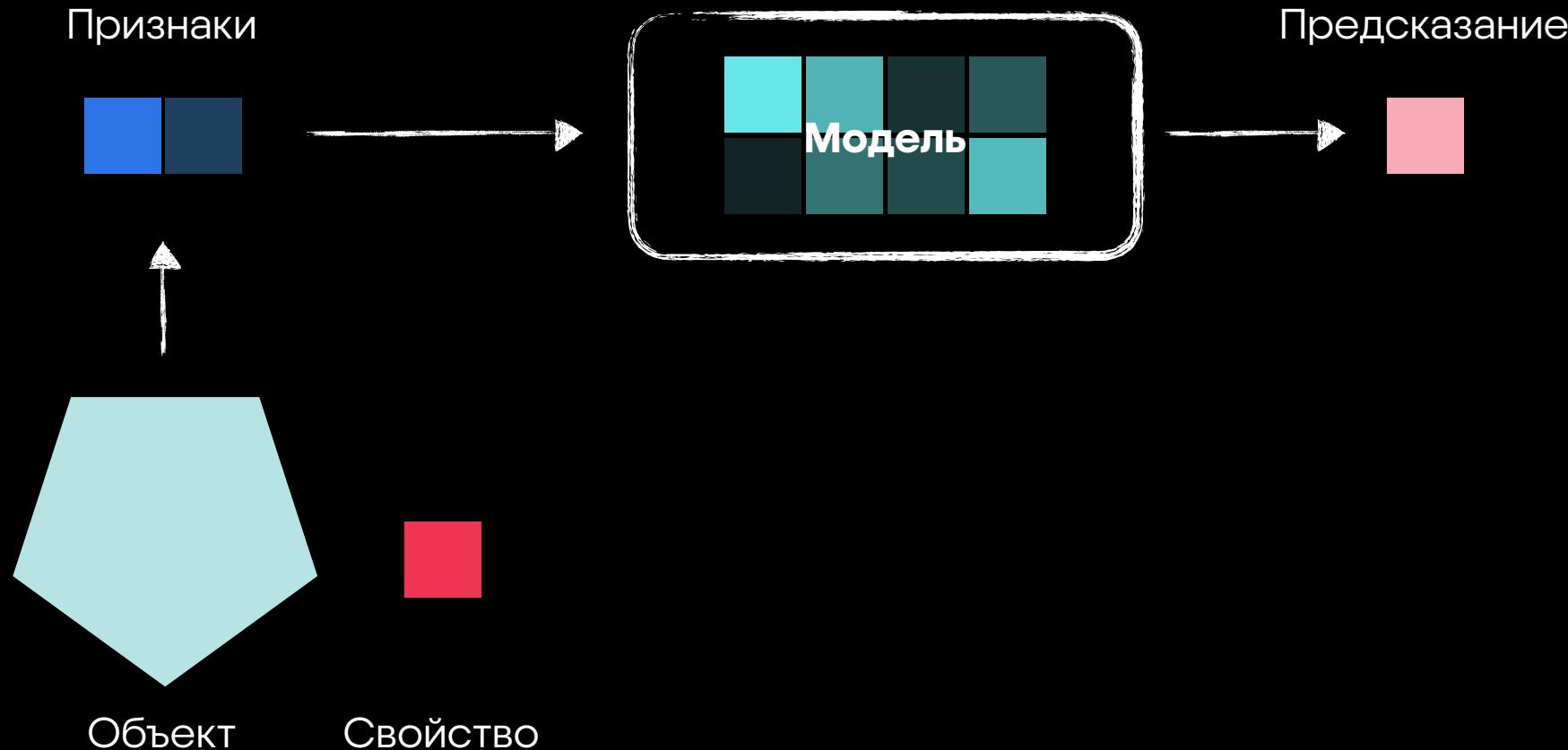




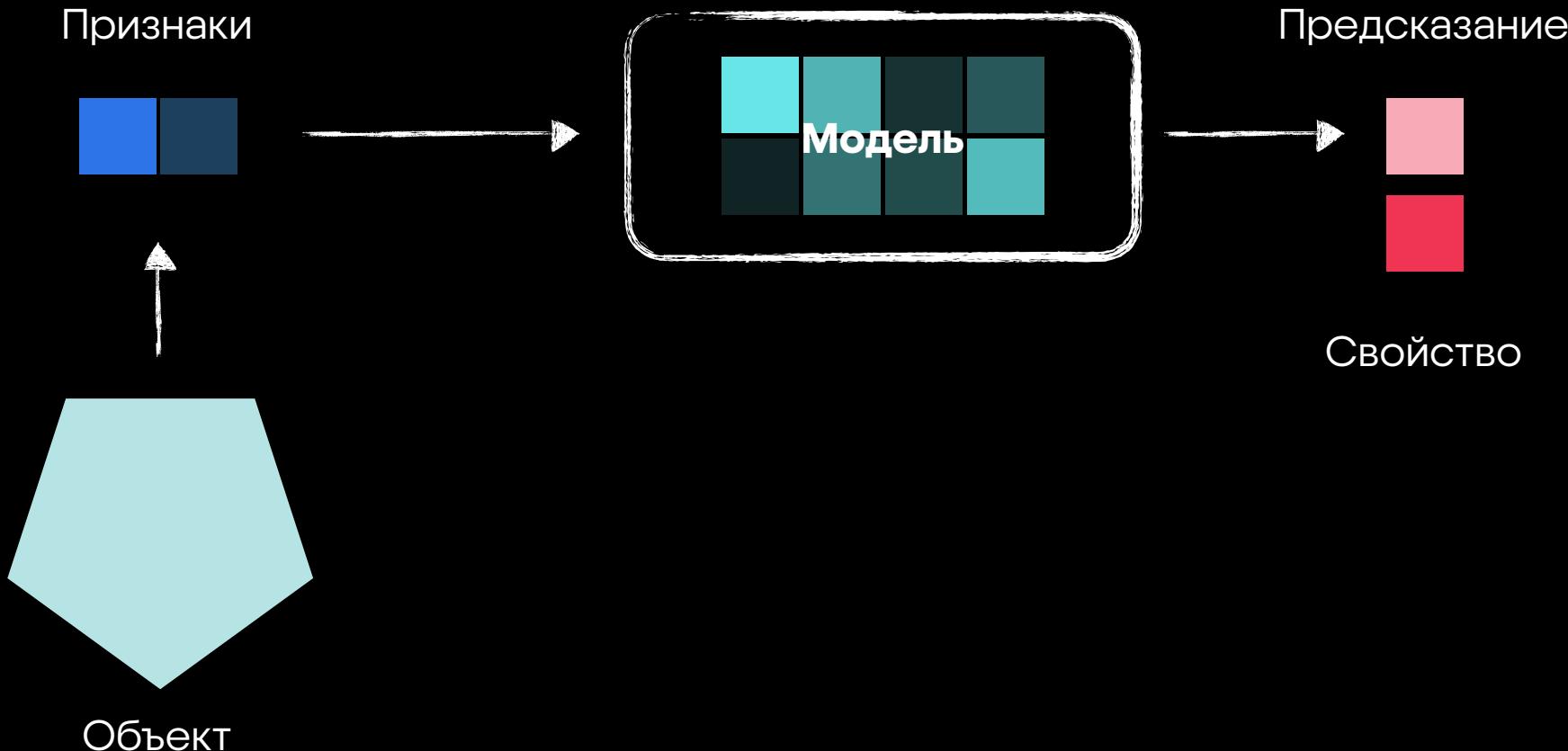
Юра Сарапулов

Введение в глубокое обучение

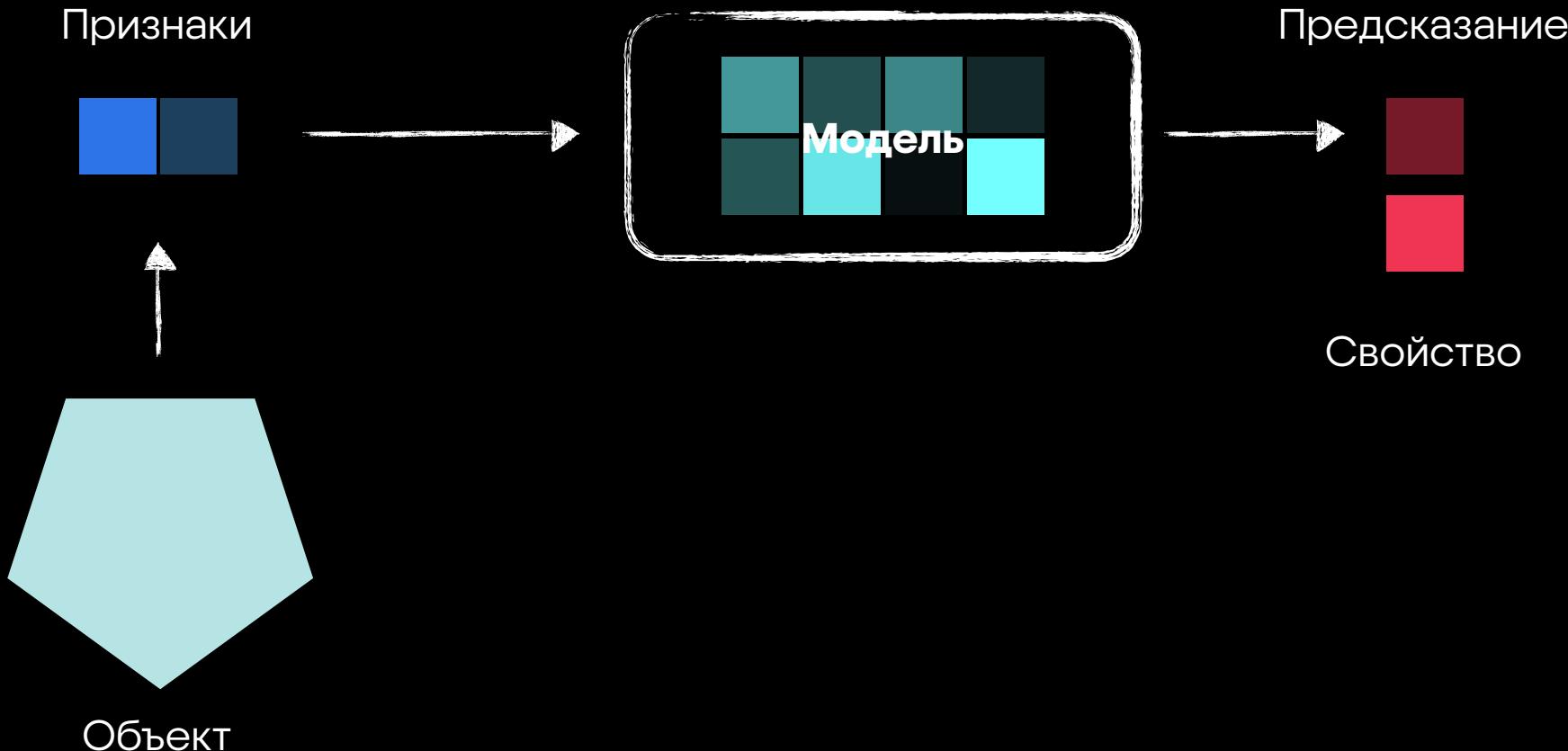
Классическое обучение с учителем



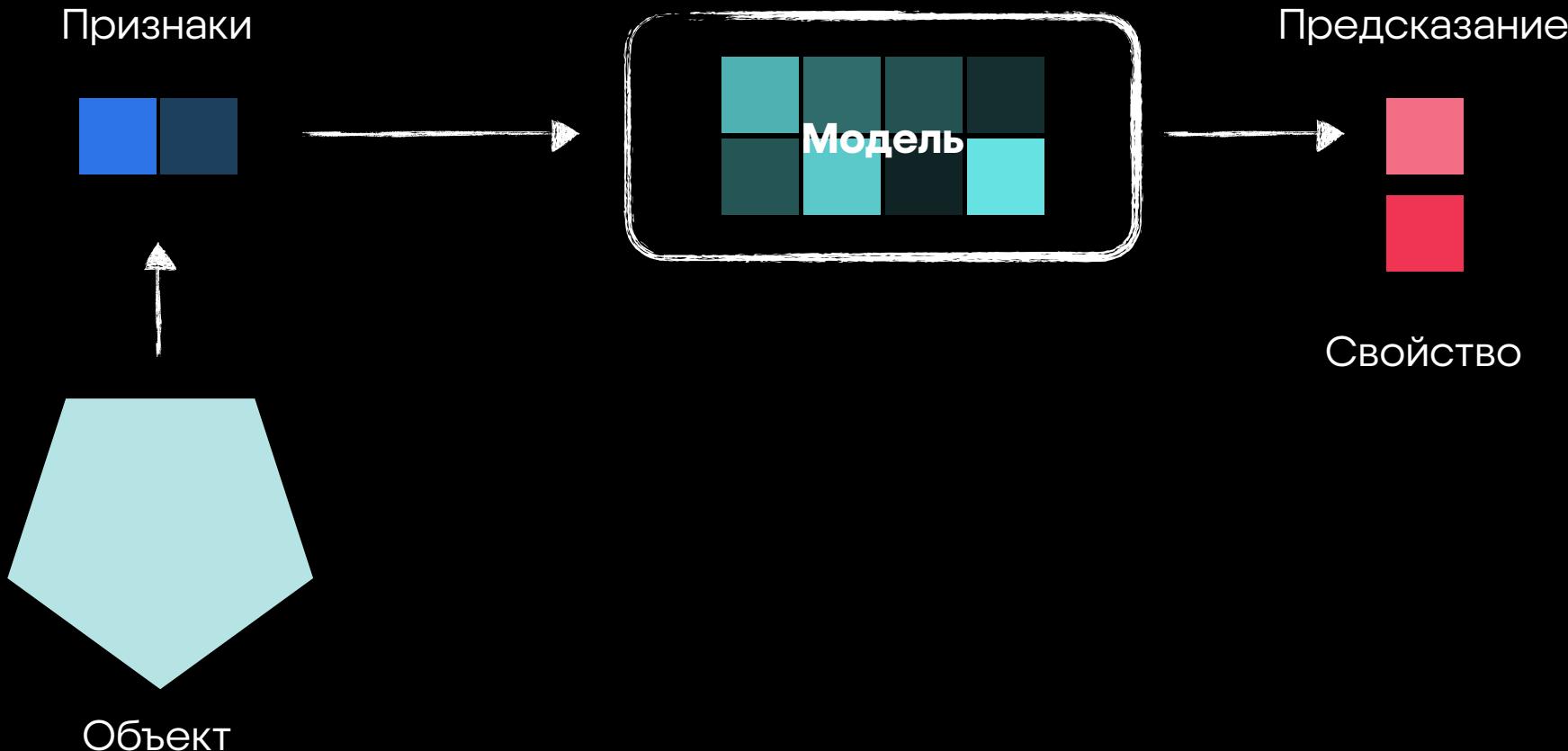
Классическое обучение с учителем



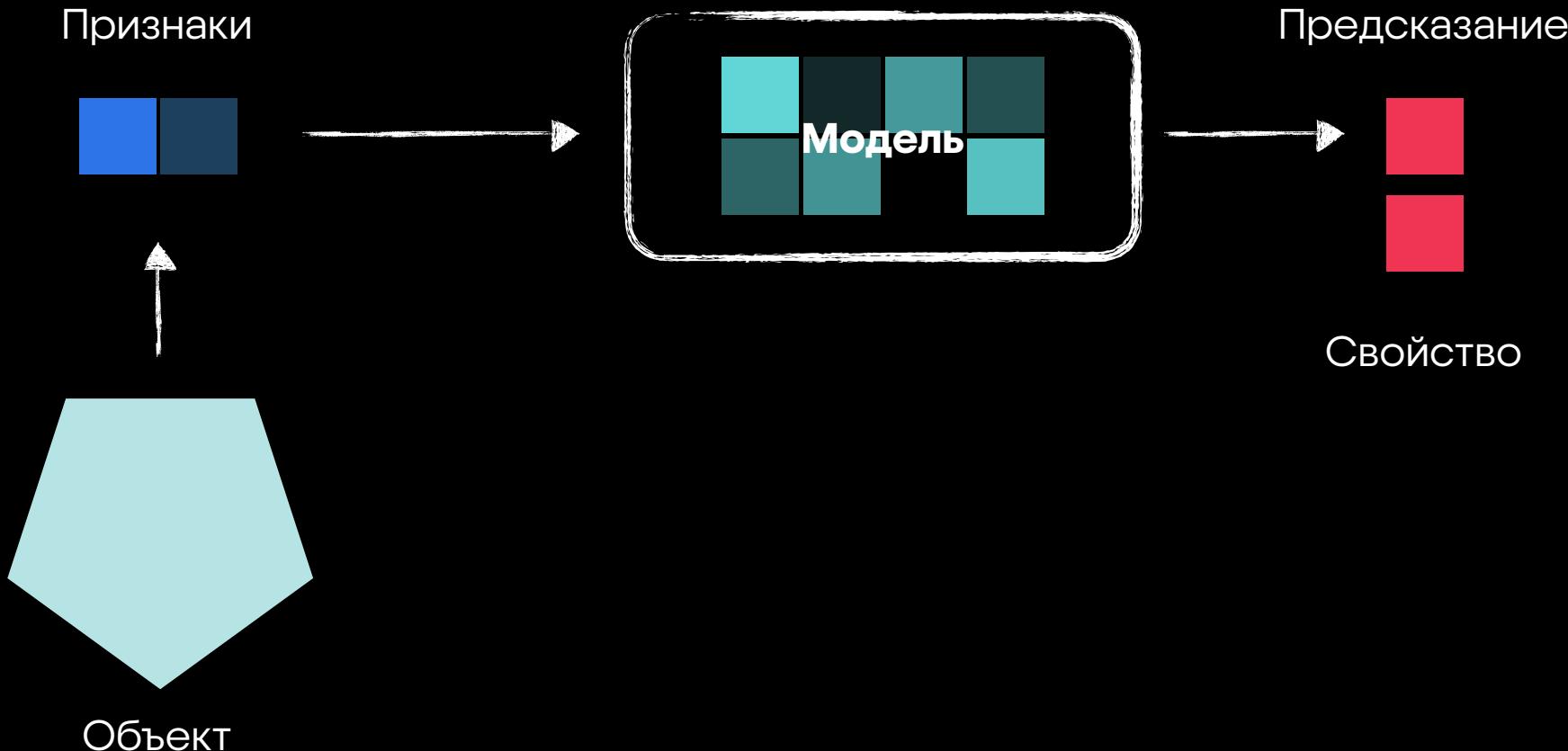
Классическое обучение с учителем



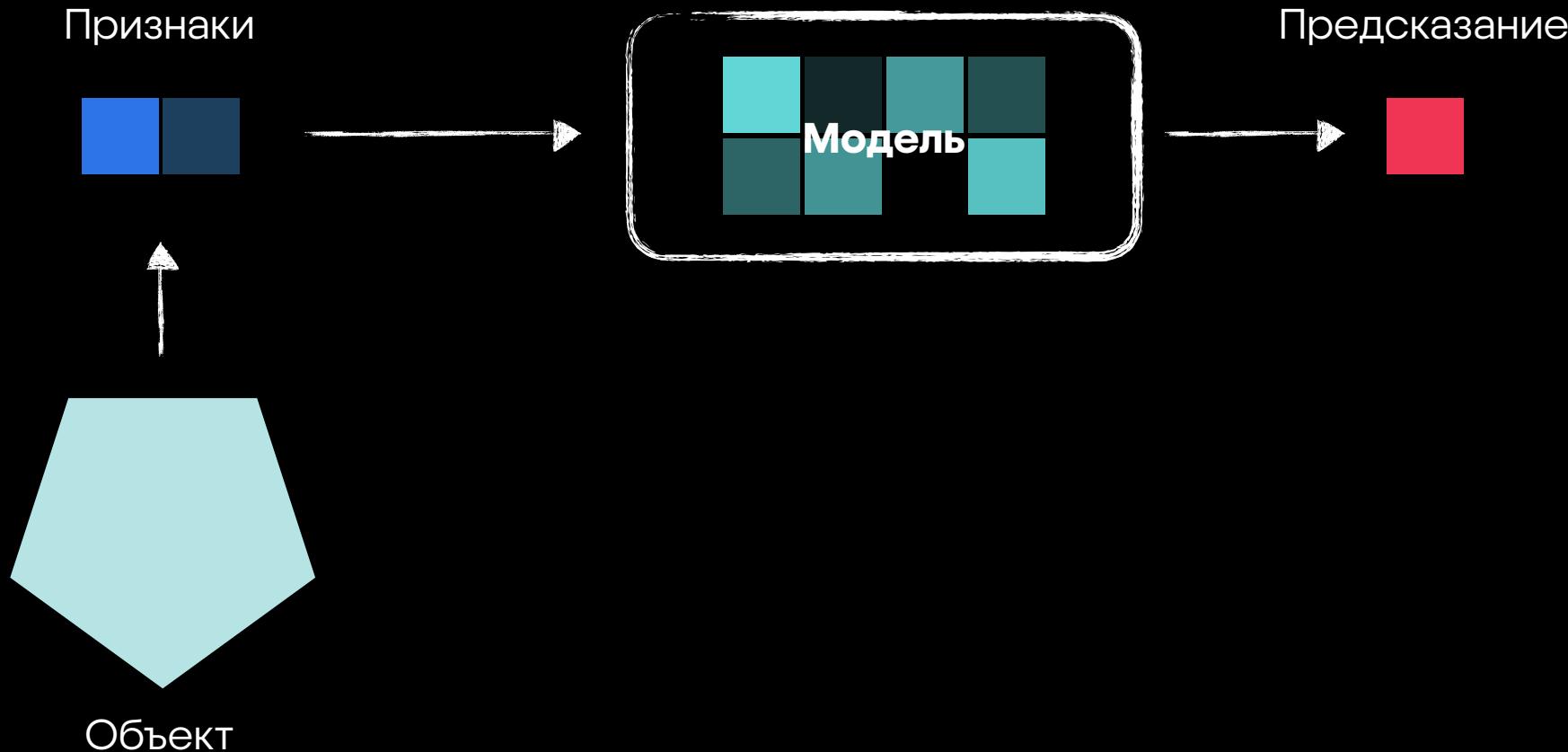
Классическое обучение с учителем



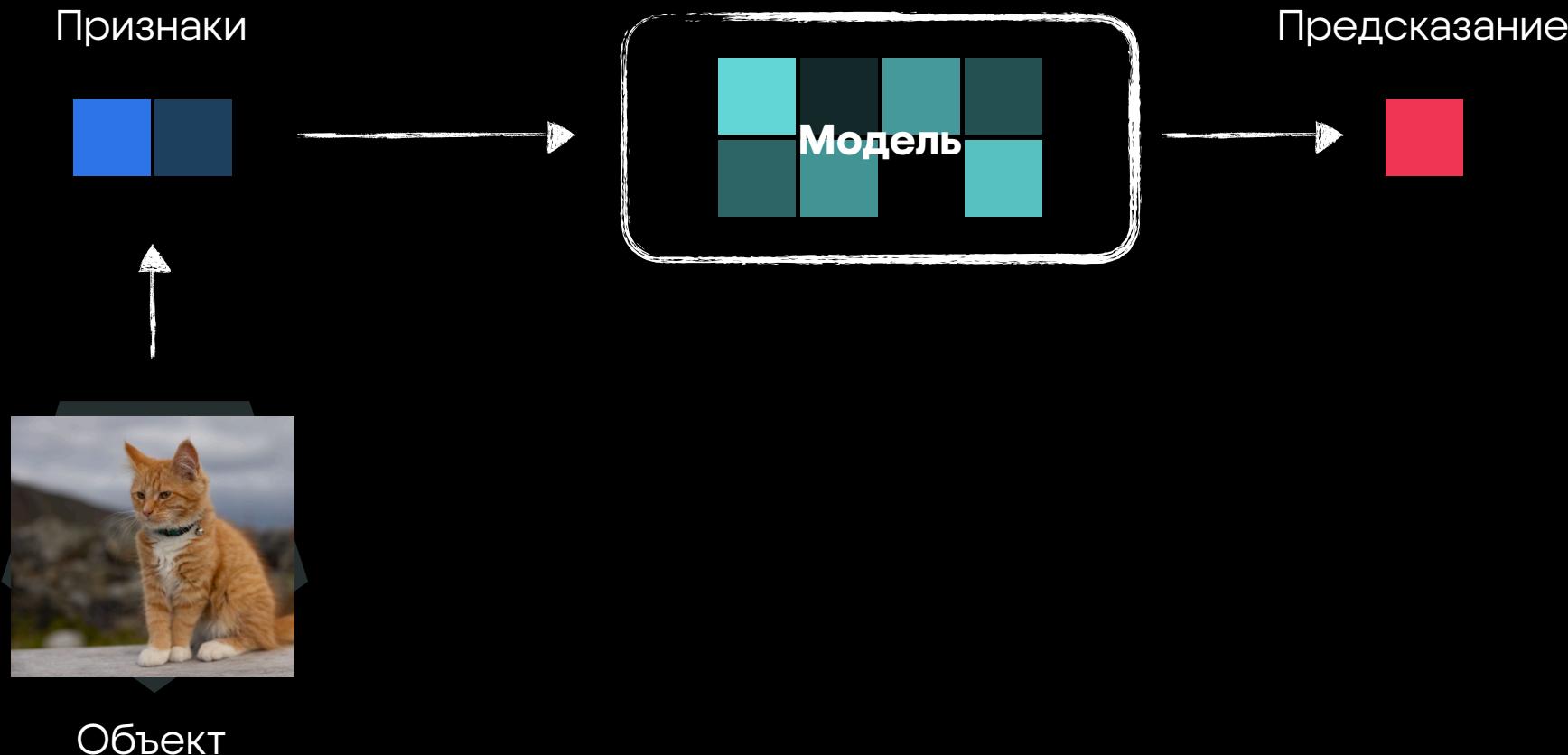
Классическое обучение с учителем



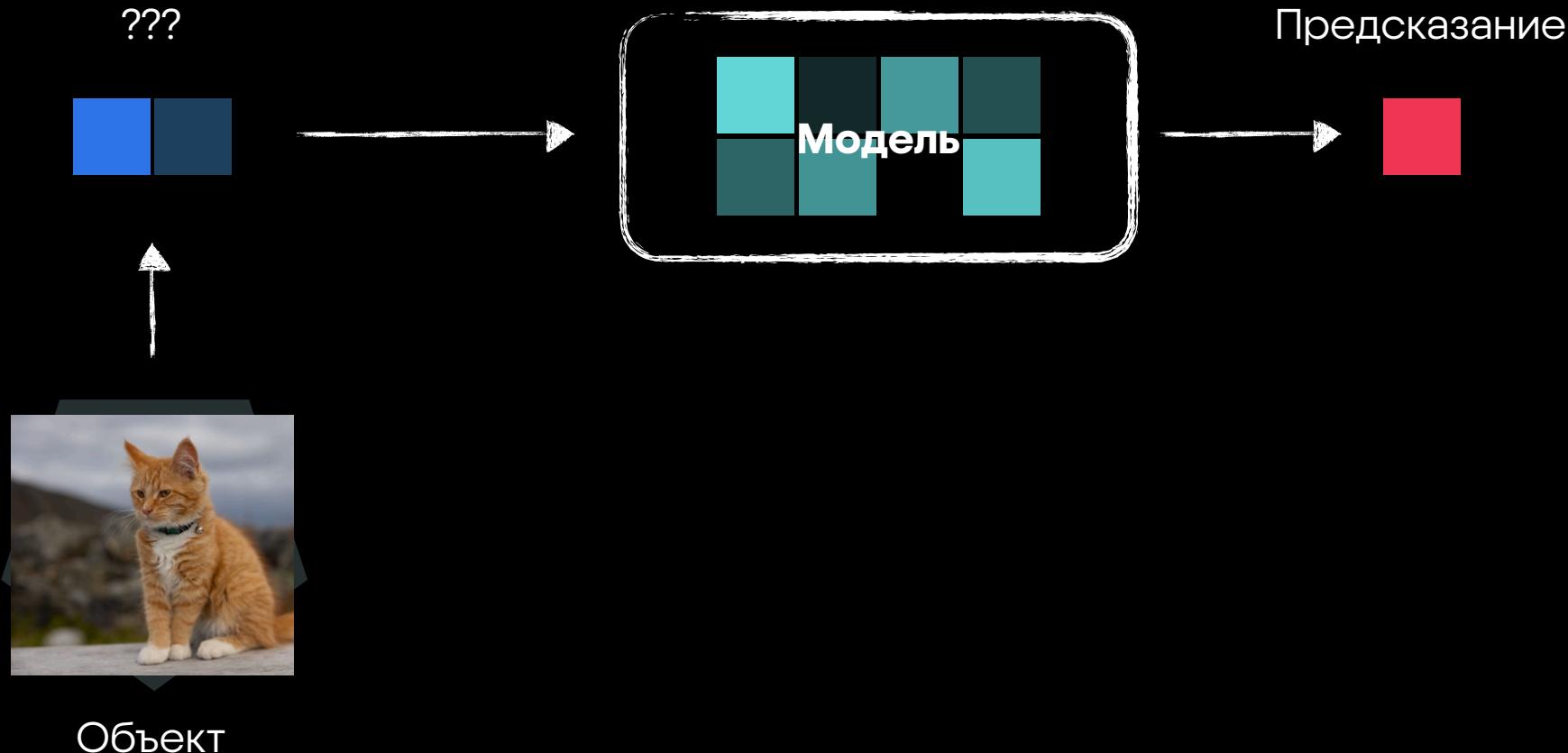
Сложные объекты



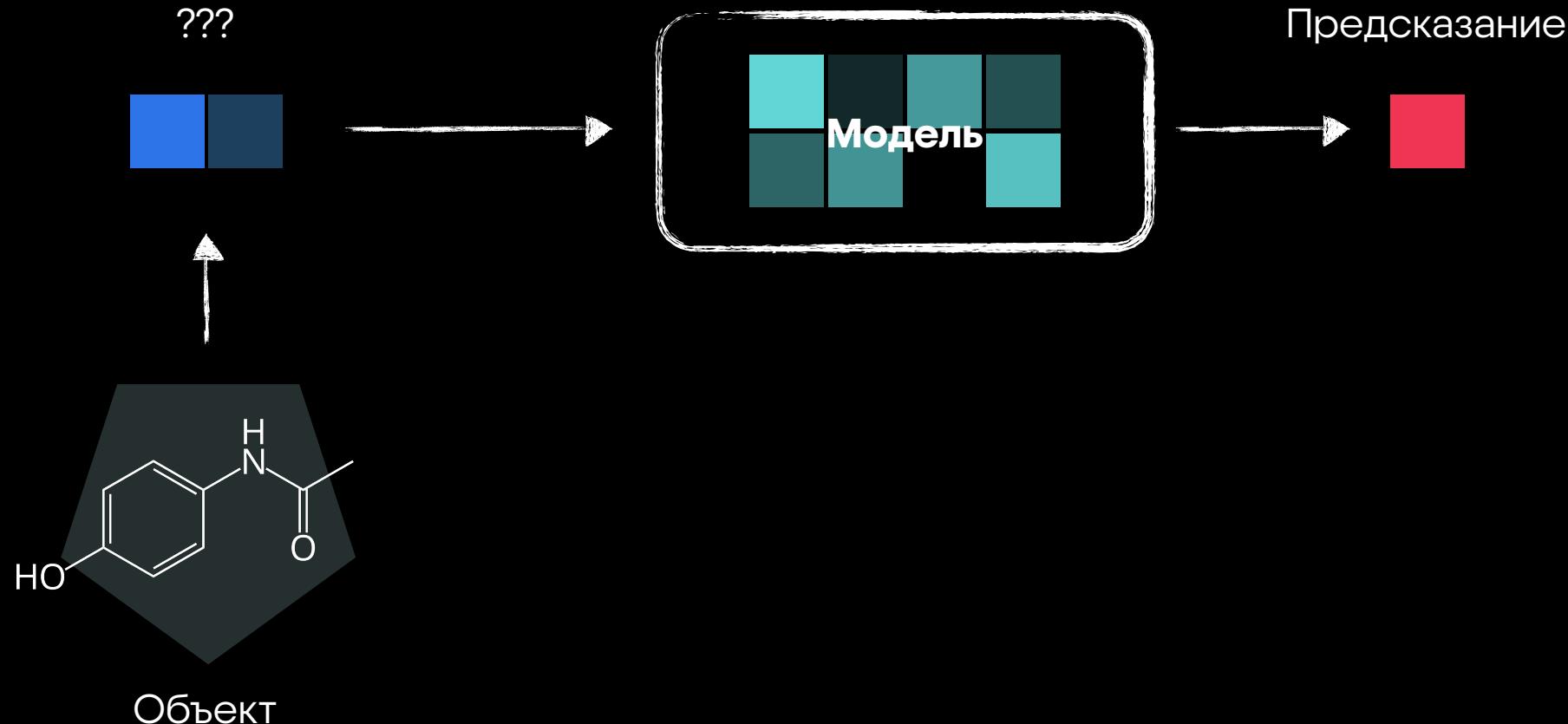
Сложные объекты



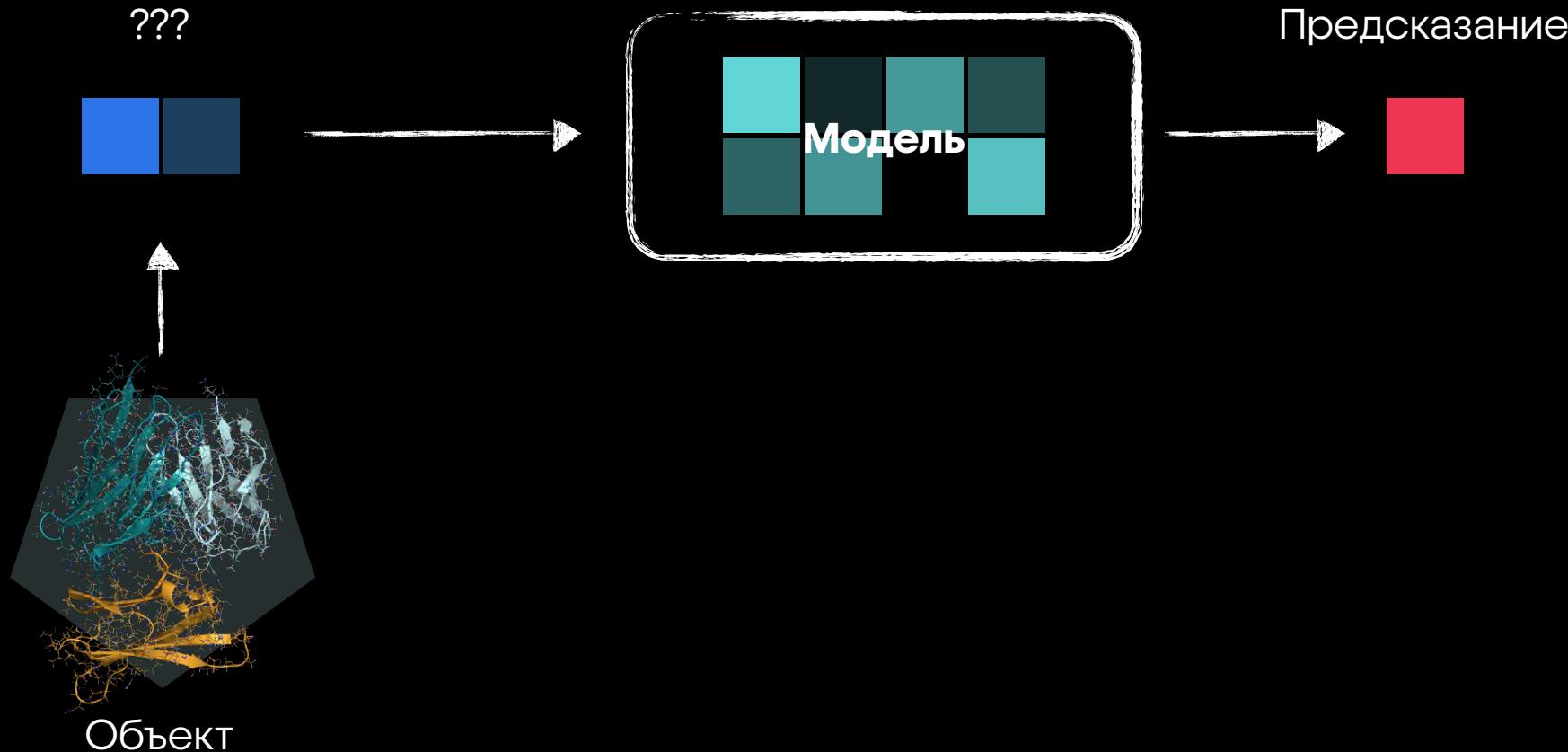
Сложные объекты



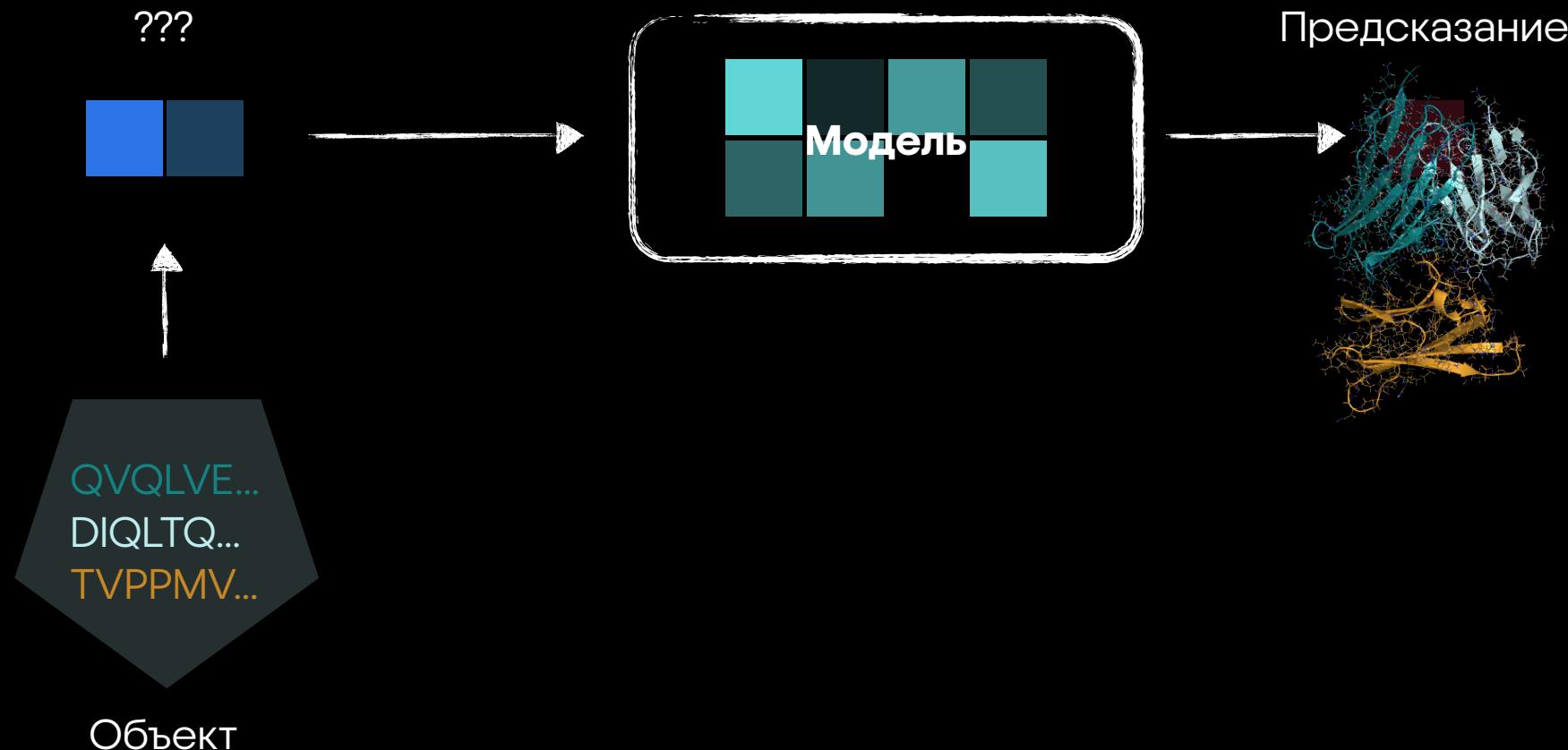
Сложные объекты



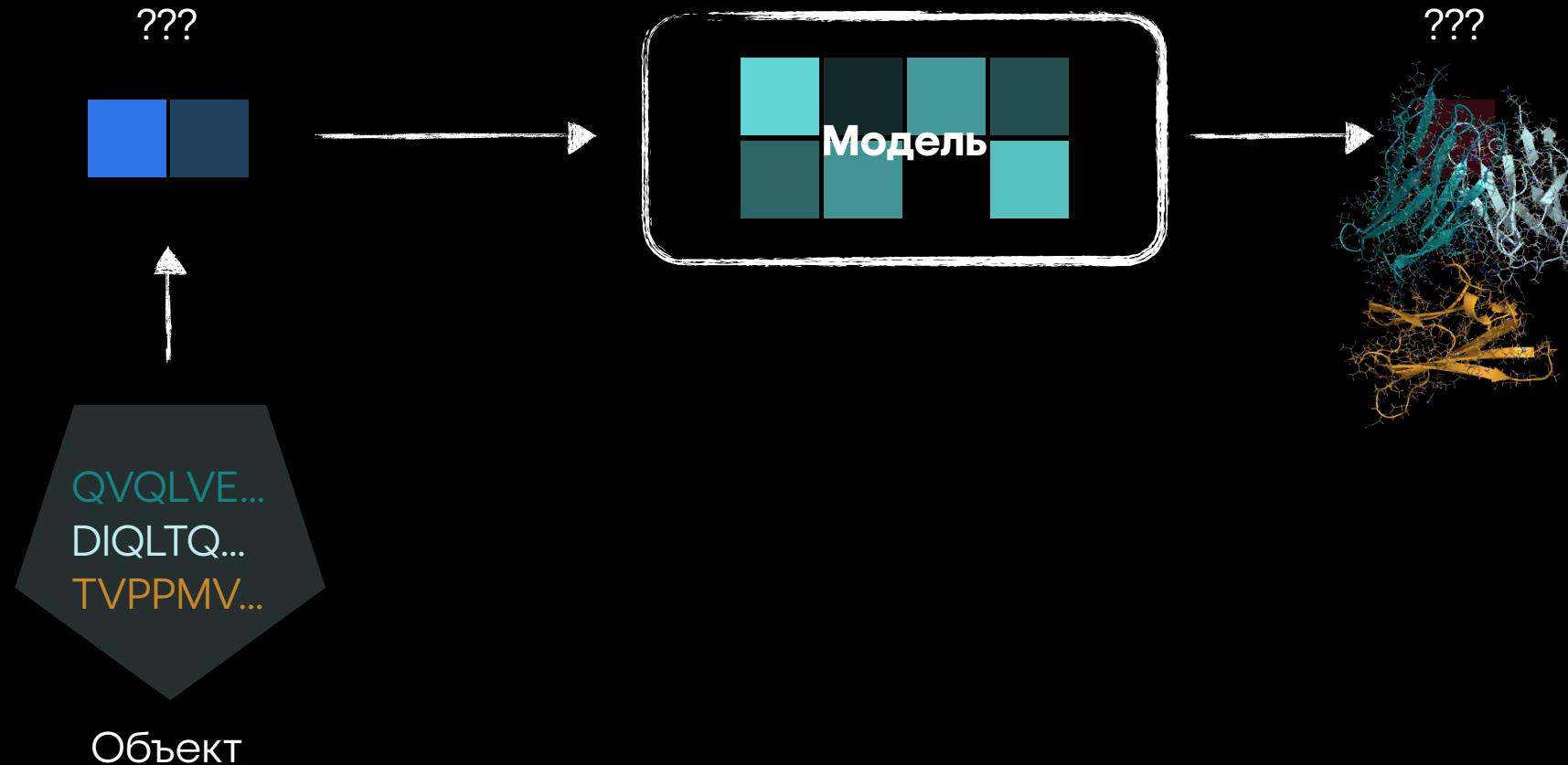
Сложные объекты



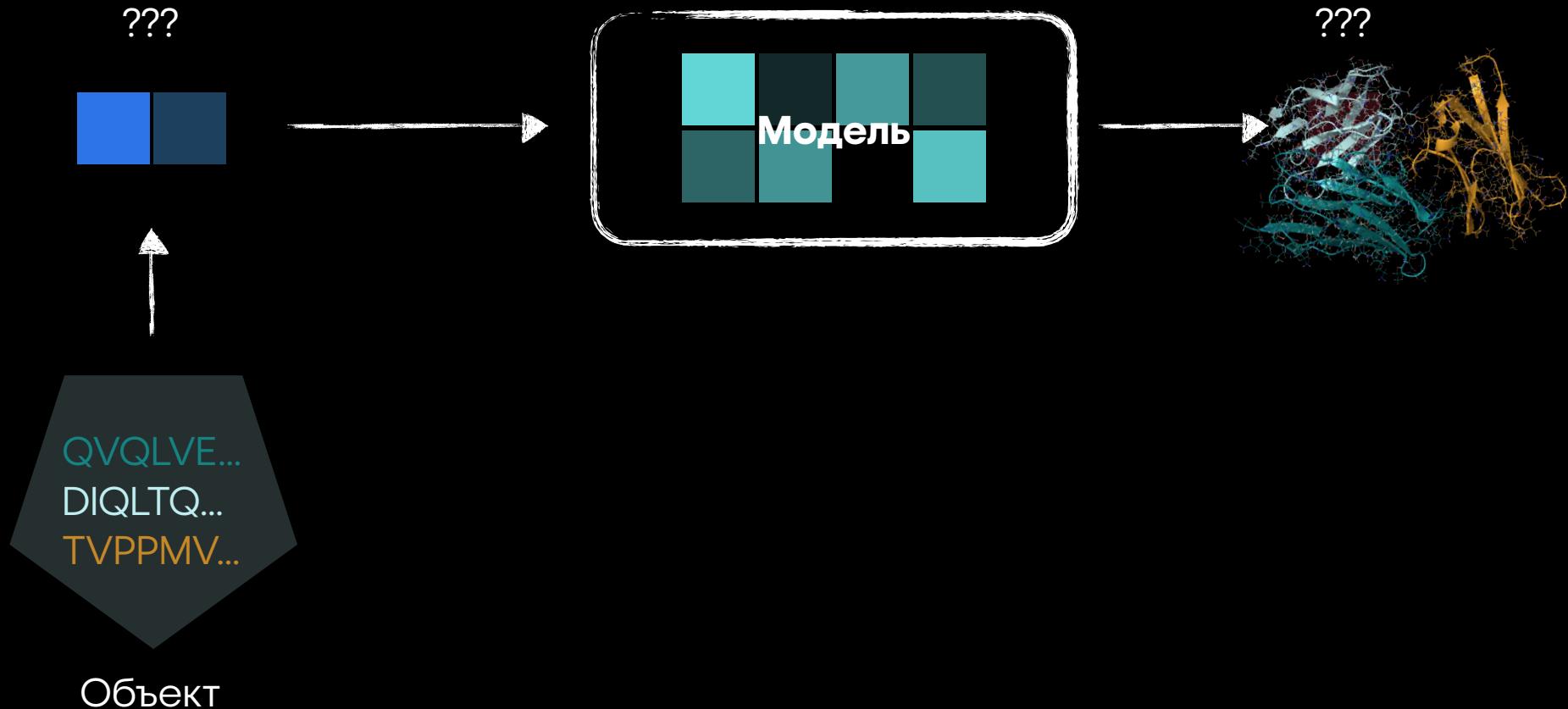
Сложные объекты



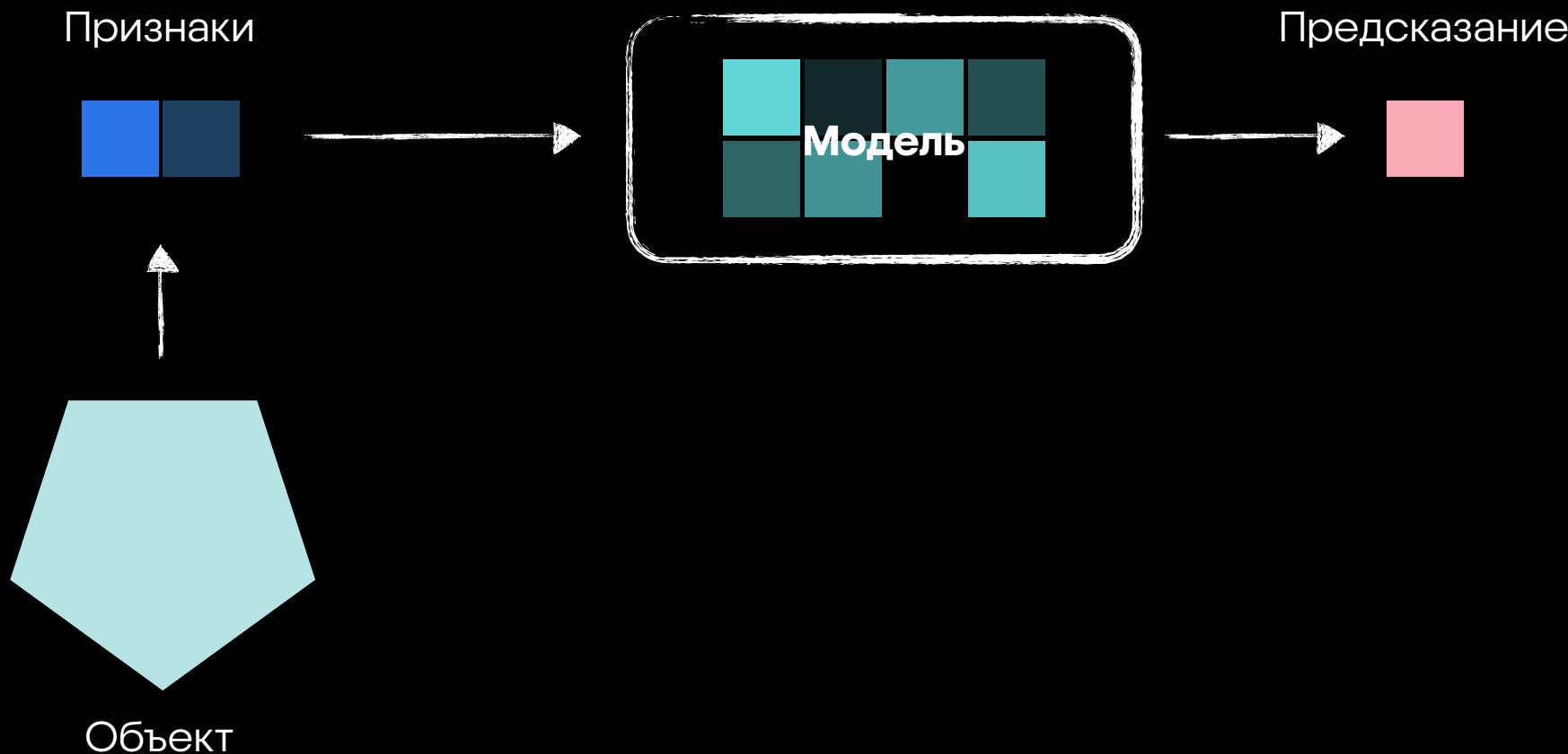
Сложные объекты



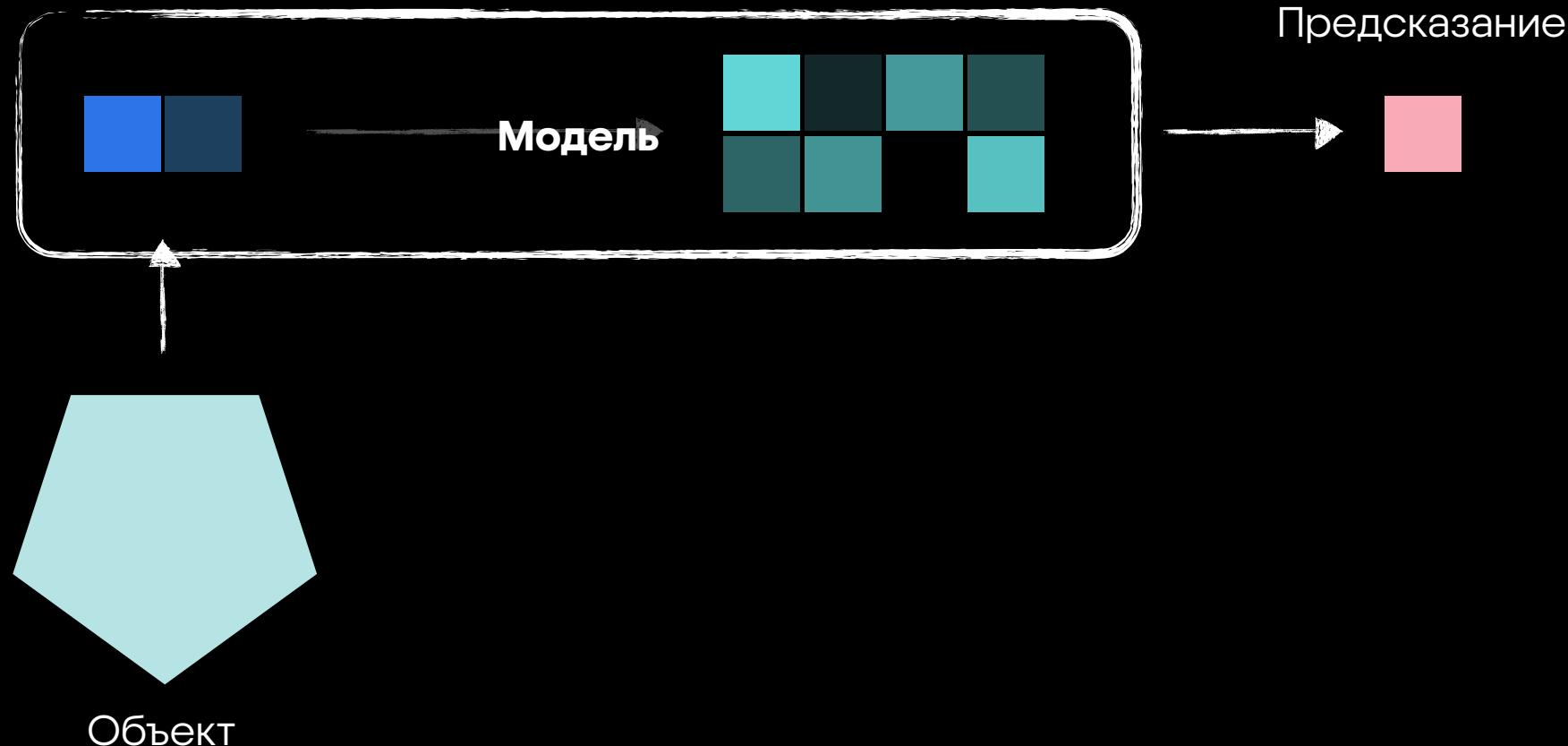
Сложные объекты



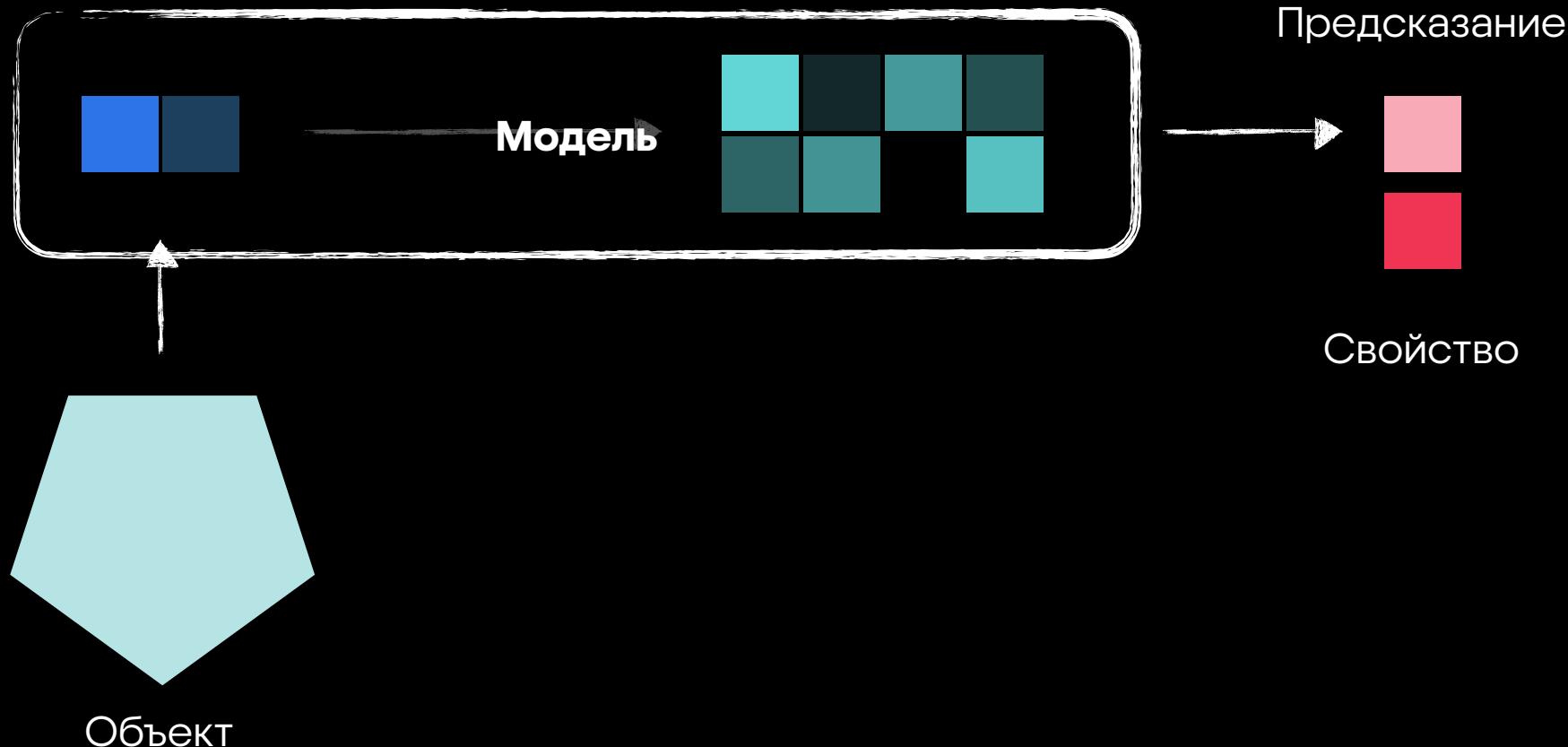
Сложные объекты



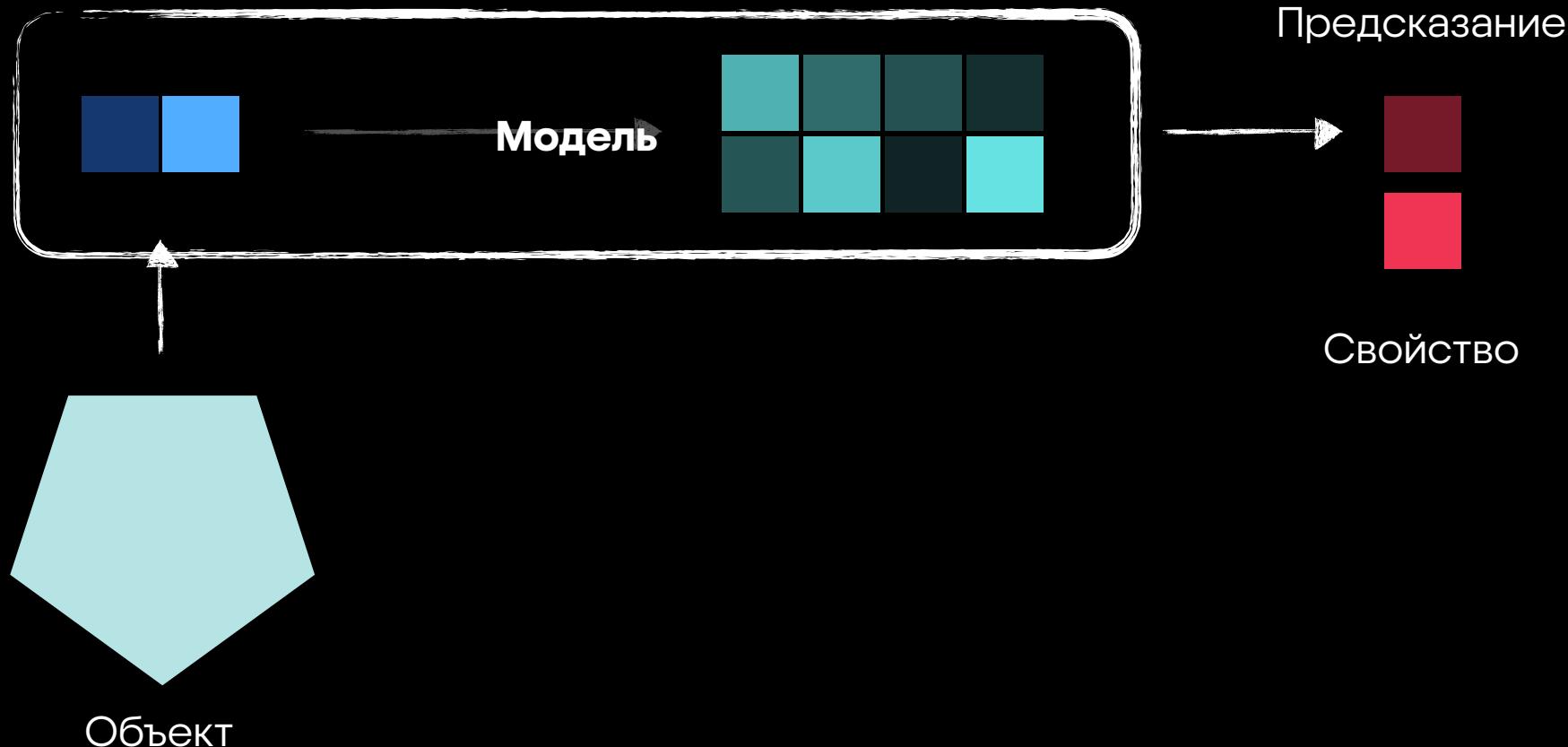
Обучение представлений данных



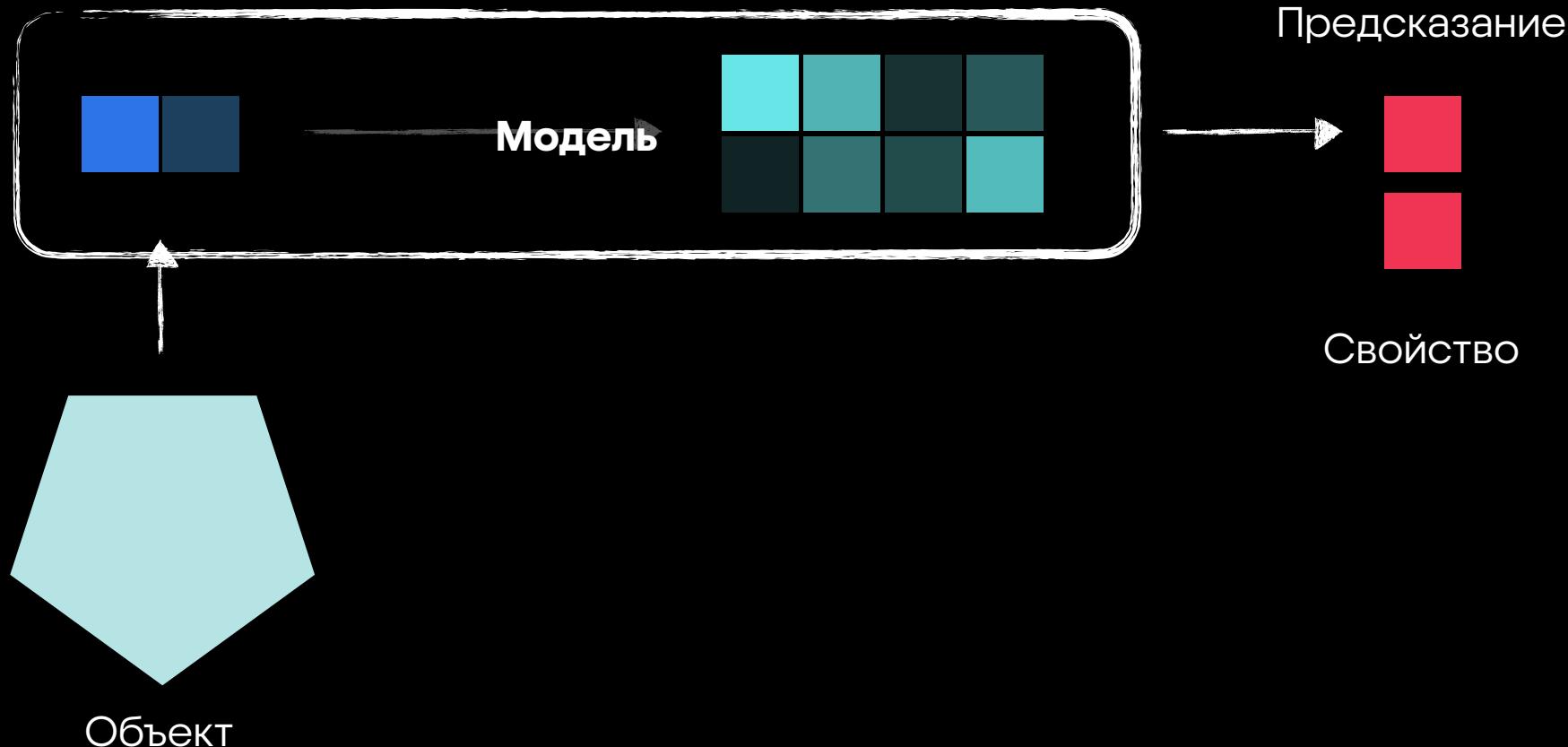
Обучение представлений данных



Обучение представлений данных



Обучение представлений данных



Обучение представлений данных



Изображения

Обучение представлений данных



QVQLVE...
DIQLTQ...
TVPPMV...

Последовательности



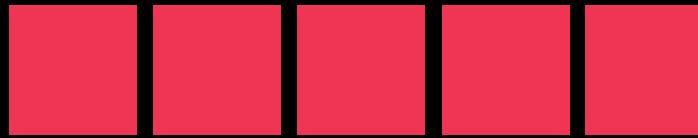
Изображения

Обучение представлений данных



Как обычно проходят глубокое обучение

Основы нейронных сетей



Свёрточные сети
в компьютерном зрении



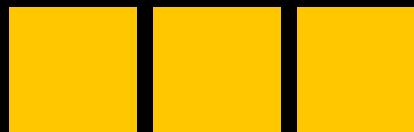
Рекуррентные сети и трансформеры
для последовательностей



Графовые сети

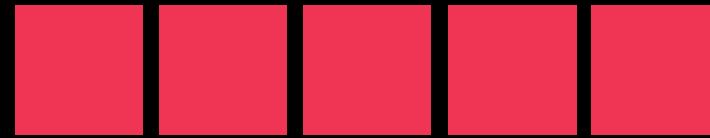


Генеративные модели
(GAN, VAE, DDPM)



Как будем проходить мы

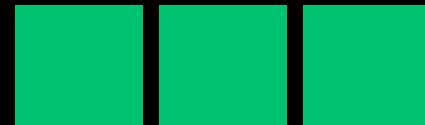
Основы нейронных сетей



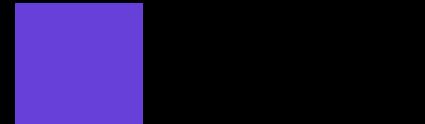
Свёрточные сети
в компьютерном зрении



Рекуррентные сети и трансформеры
для последовательностей



Графовые сети



Генеративные модели
(GAN, VAE, DDPM)



Сколько времени у нас

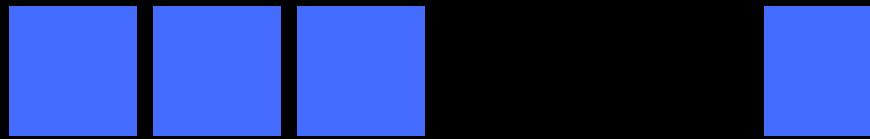


Как будем проходить мы

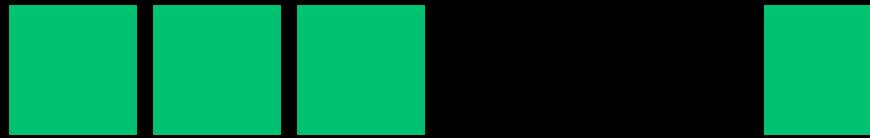
Основы нейронных сетей



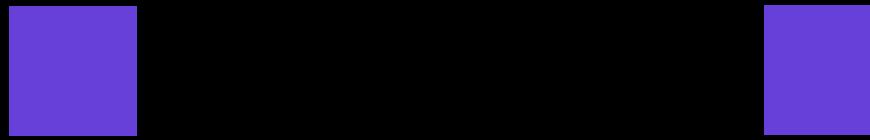
Свёрточные сети
в компьютерном зрении



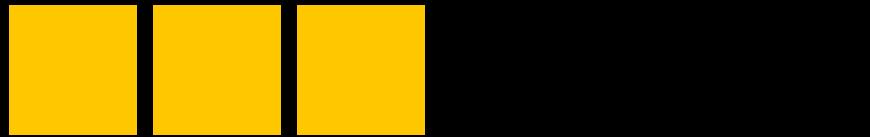
Рекуррентные сети и трансформеры
для последовательностей



Графовые сети



Генеративные модели
(GAN, VAE, DDPM)

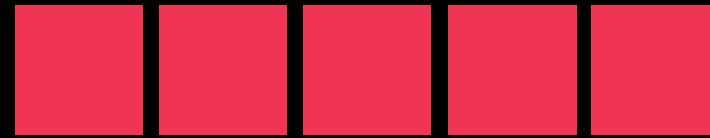


Сколько времени у нас



Как будем проходить мы

Основы нейронных сетей



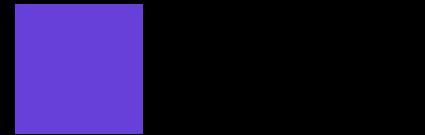
Свёрточные сети
в компьютерном зрении



Рекуррентные сети и трансформеры
для последовательностей



Графовые сети



Генеративные модели
(GAN, VAE, DDPM)



Сколько времени у нас

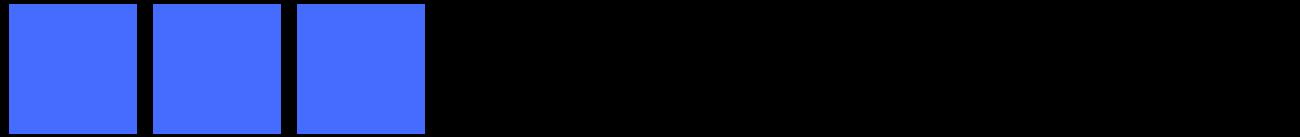


Как будем проходить мы

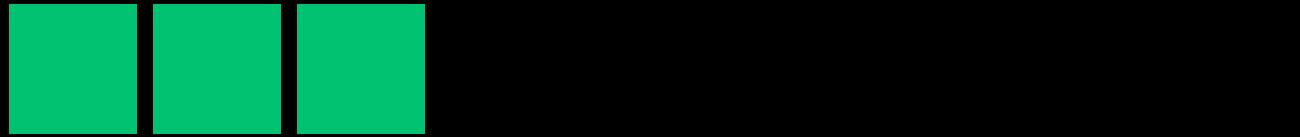
Основы нейронных сетей



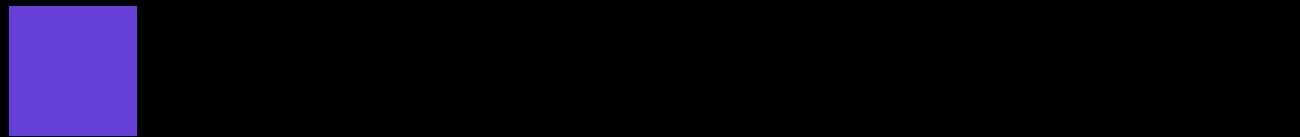
Свёрточные сети
в компьютерном зрении



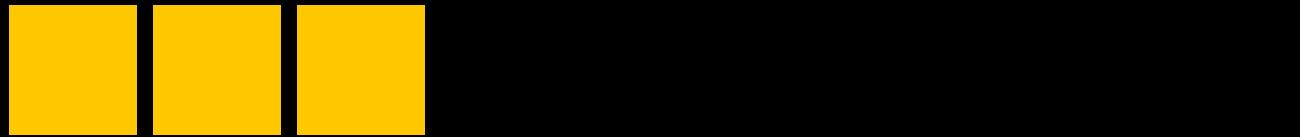
Рекуррентные сети и трансформеры
для последовательностей



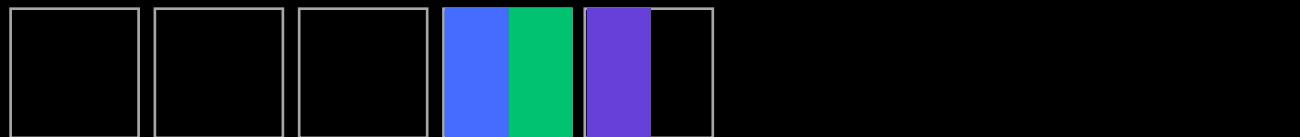
Графовые сети



Генеративные модели
(GAN, VAE, DDPM)



Сколько времени у нас



Как будем проходить мы

Основы нейронных сетей



Свёрточные сети
в компьютерном зрении



Рекуррентные сети и трансформеры
для последовательностей



Графовые сети



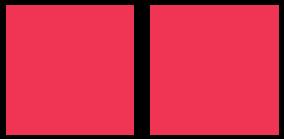
Генеративные модели
(GAN, VAE, DDPM)



Сколько времени у нас

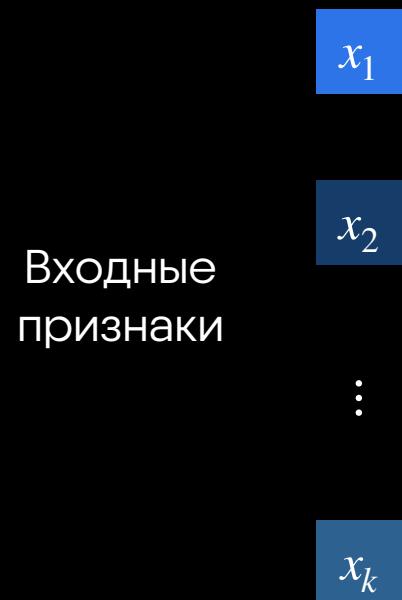


План на сегодня

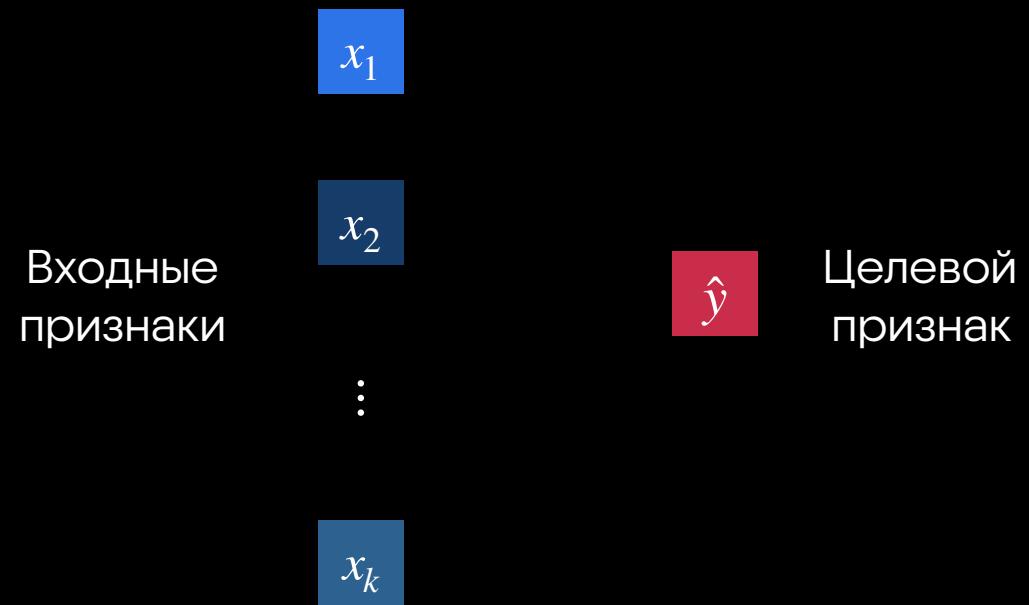


1. Нейронные сети: теоретический минимум
2. Основы Pytorch и первая нейронная сеть
3. Некоторые тонкости обучения нейронных сетей

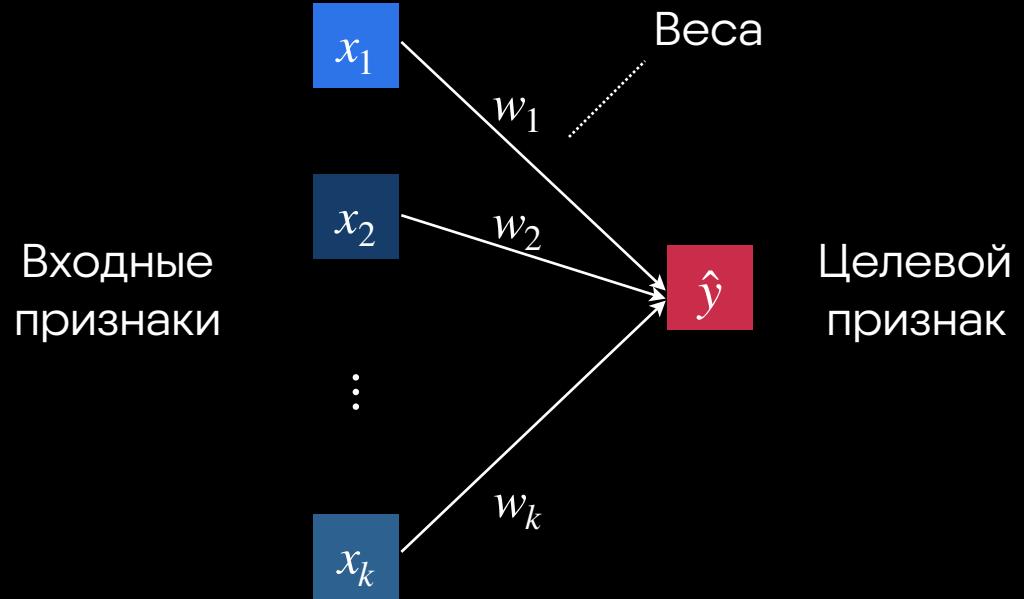
От линейной регрессии к модели нейрона



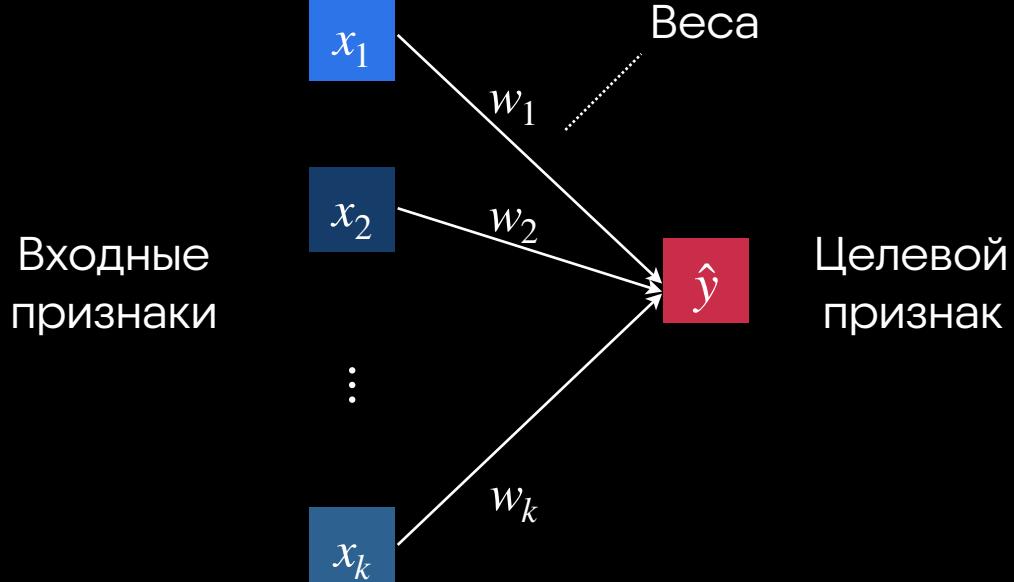
От линейной регрессии к модели нейрона



От линейной регрессии к модели нейрона



От линейной регрессии к модели нейрона

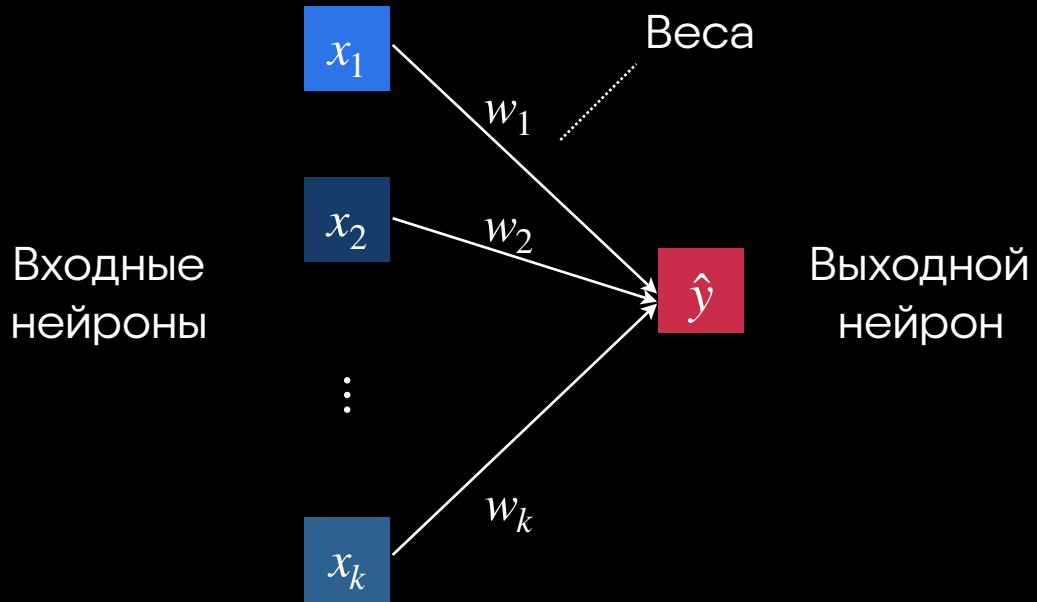


$$\hat{y} = \sum_{i=1}^k w_i x_i$$

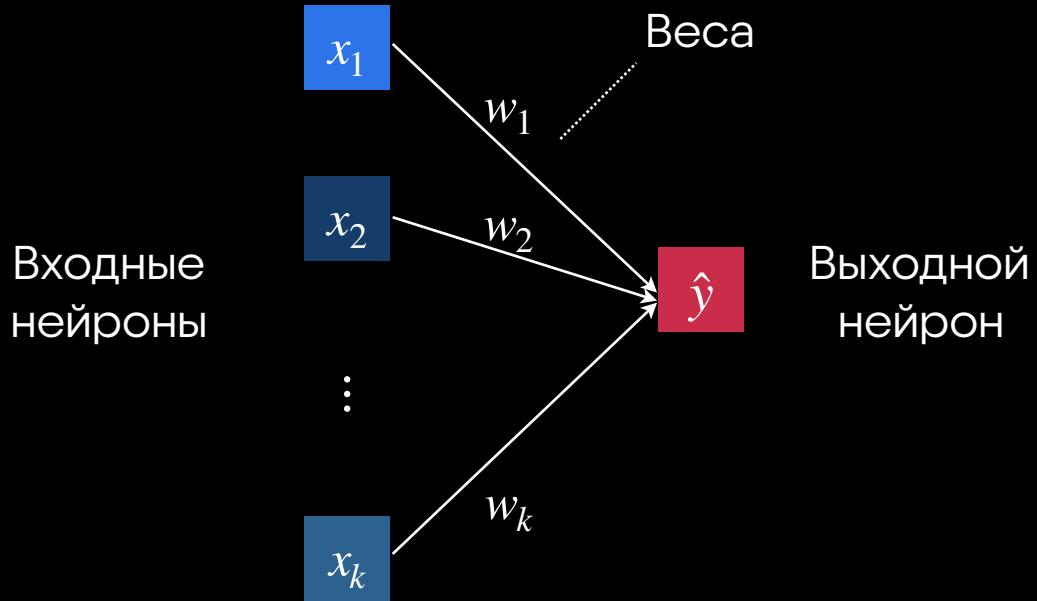
В матричной форме:

$$\hat{y} = \mathbf{w}^T \mathbf{x}$$

От линейной регрессии к модели нейрона

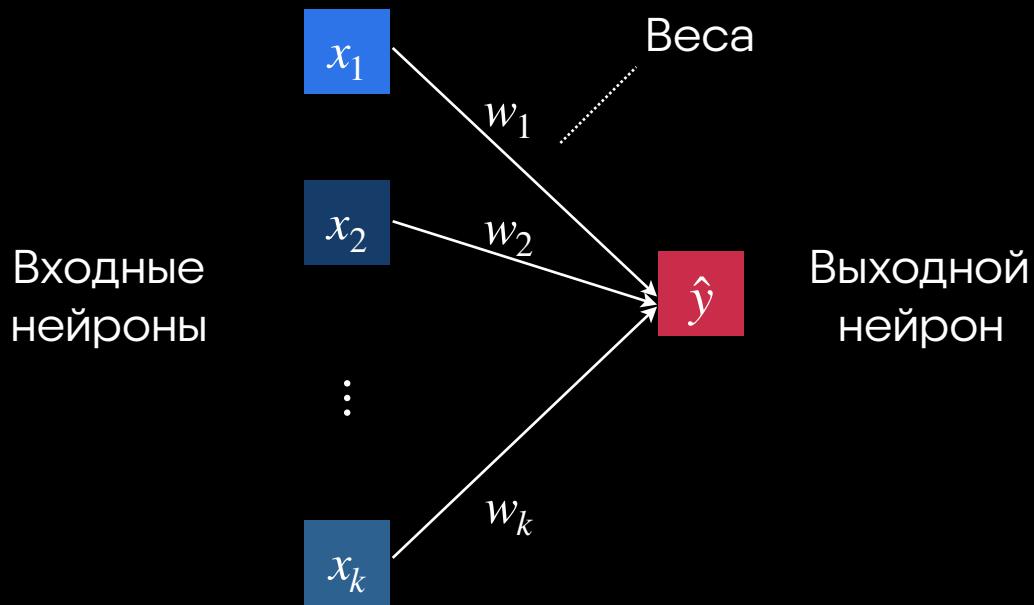


От линейной регрессии к модели нейрона



$$a = \sum_{i=1}^k w_i x_i \quad \dots \quad \text{Пре-активация}$$

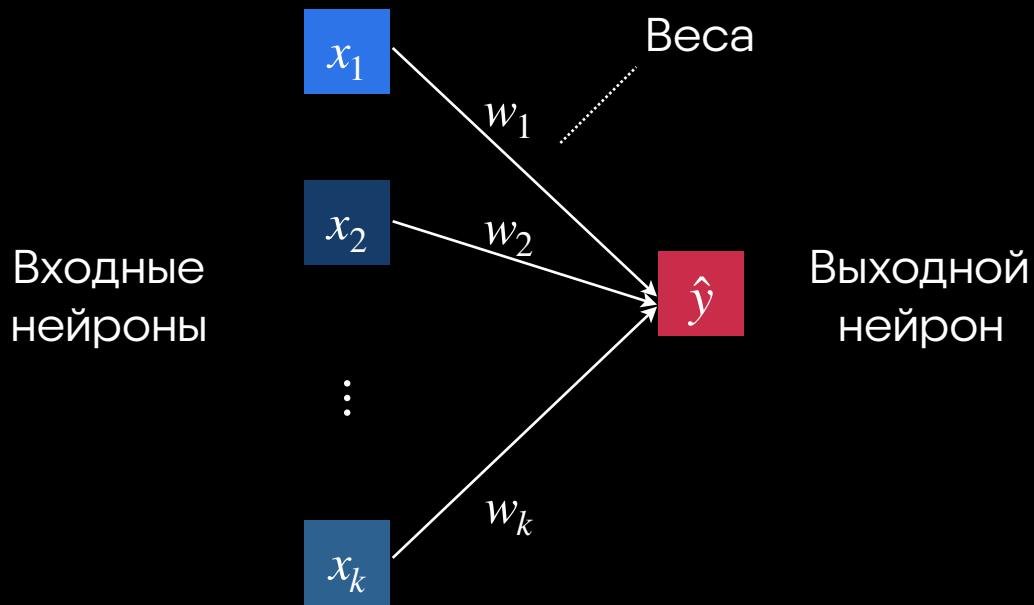
От линейной регрессии к модели нейрона



$$a = \sum_{i=1}^k w_i x_i \quad \cdots \quad \text{Пре-активация}$$
$$\hat{y} = f(a) \quad \cdots \quad \text{Активация}$$

↓
Функция активации

От линейной регрессии к модели нейрона

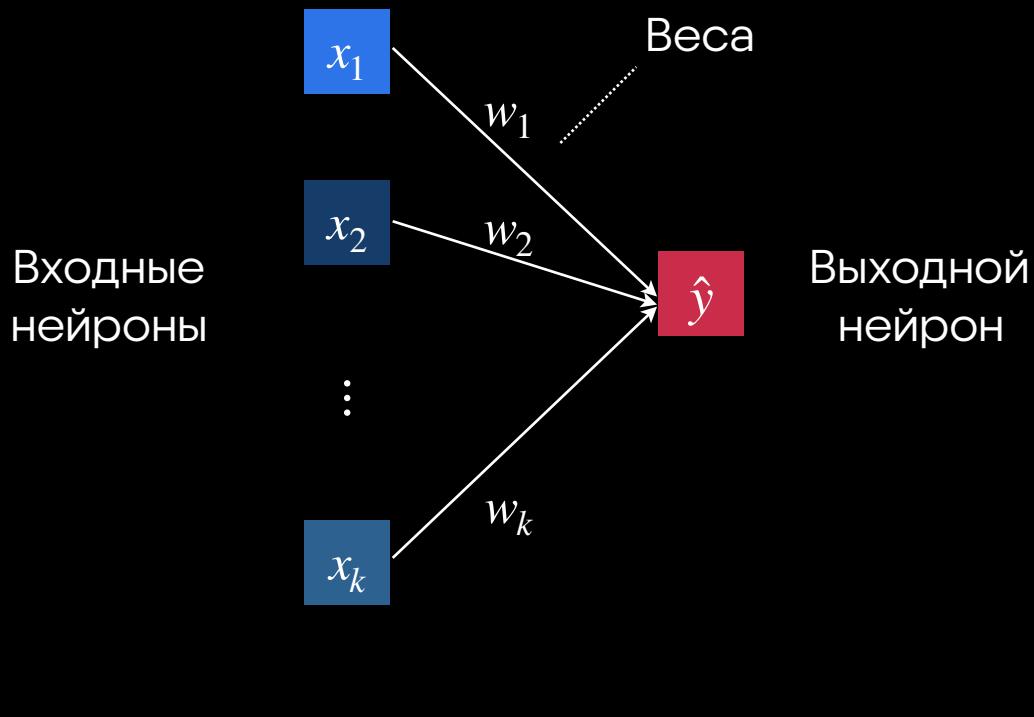


$$a = \sum_{i=1}^k w_i x_i \quad \dots \quad \text{Пре-активация}$$
$$\hat{y} = f(a) \quad \dots \quad \text{Активация}$$

↓
Функция активации

Перцептрон: $f(a) = \begin{cases} 1, & a > 0 \\ 0, & a = 0 \\ -1, & a < 0 \end{cases}$

От линейной регрессии к модели нейрона



$$a = \sum_{i=1}^k w_i x_i \quad \text{Pre-активация}$$
$$\hat{y} = f(a) \quad \text{Активация}$$

Функция активации

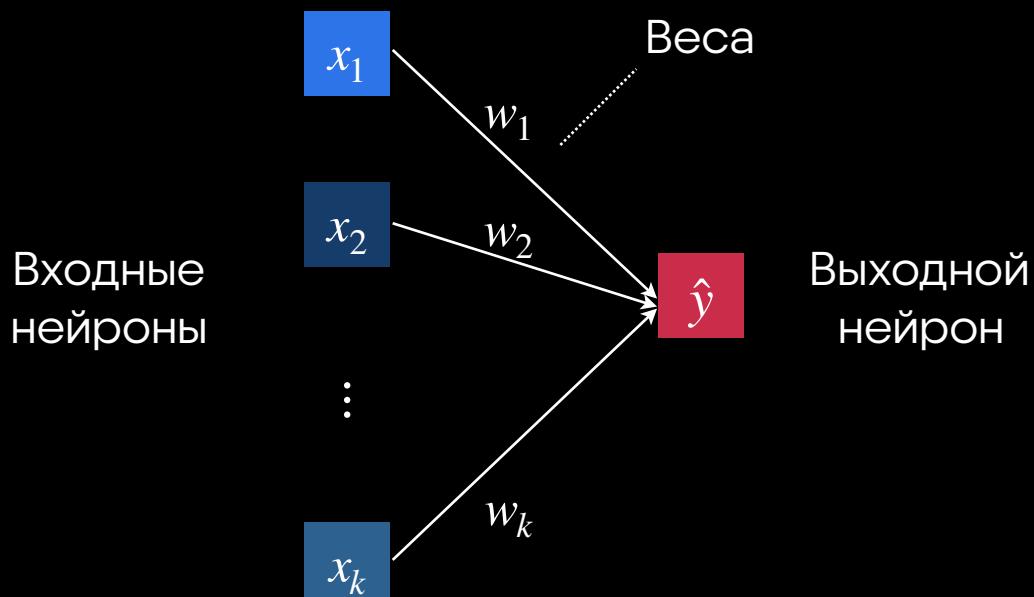
Перцептрон:

$$f(a) = \begin{cases} 1, & a > 0 \\ 0, & a = 0 \\ -1, & a < 0 \end{cases}$$

Линейная регрессия:

$$f(a) = a$$

От линейной регрессии к модели нейрона



$$a = \sum_{i=1}^k w_i x_i \quad \dots \quad \text{Пре-активация}$$
$$\hat{y} = f(a) \quad \dots \quad \text{Активация}$$

↓
Функция активации

Перцептрон:

$$f(a) = \begin{cases} 1, & a > 0 \\ 0, & a = 0 \\ -1, & a < 0 \end{cases}$$

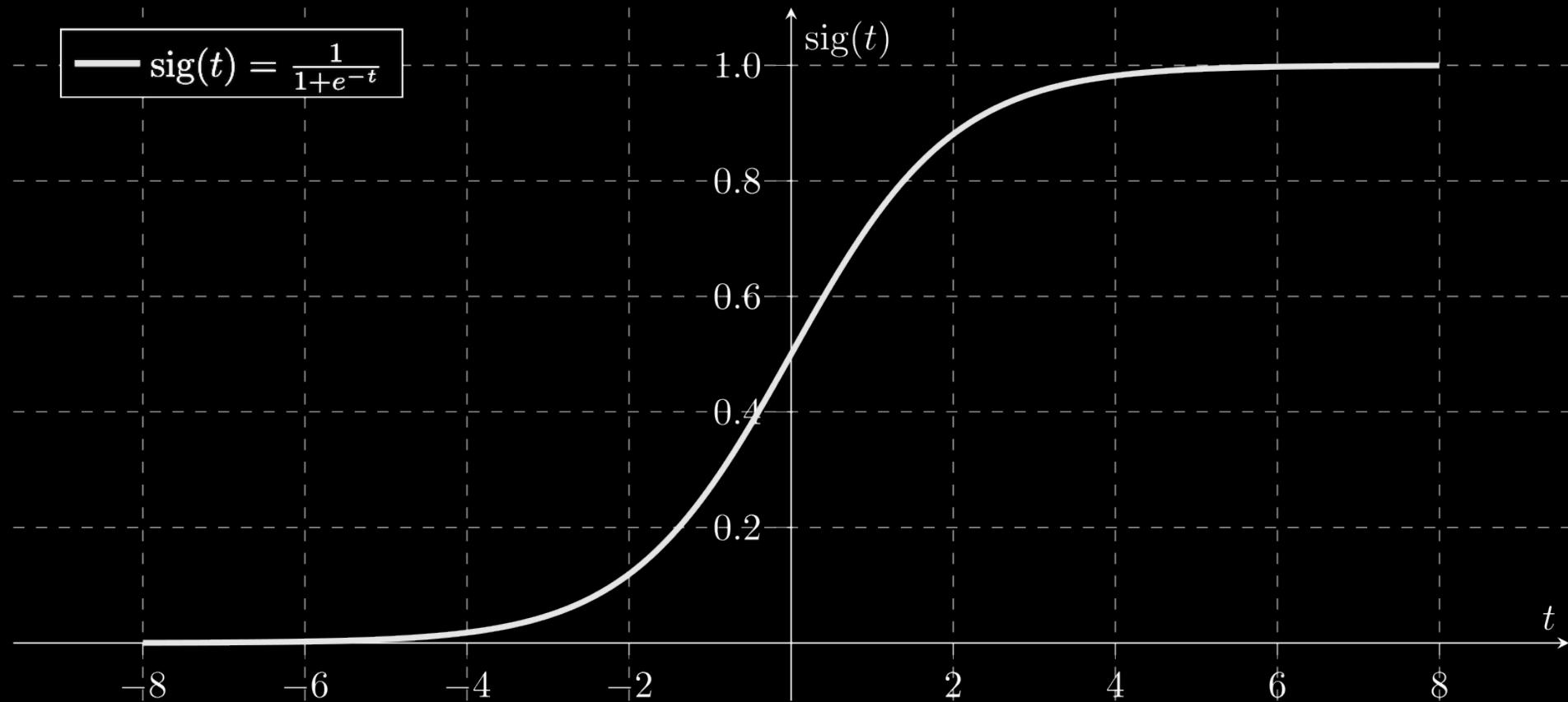
Линейная
регрессия:

$$f(a) = a$$

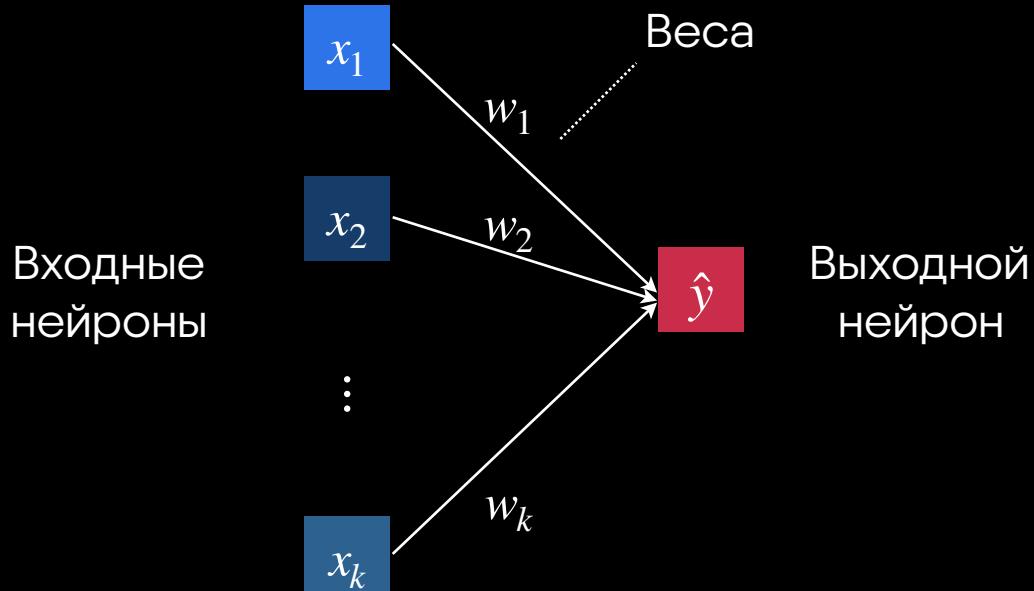
Логистическая
регрессия:

$$f(a) = \sigma(a) = \frac{1}{1 + \exp(-a)}$$

От линейной регрессии к модели нейрона



От линейной регрессии к модели нейрона



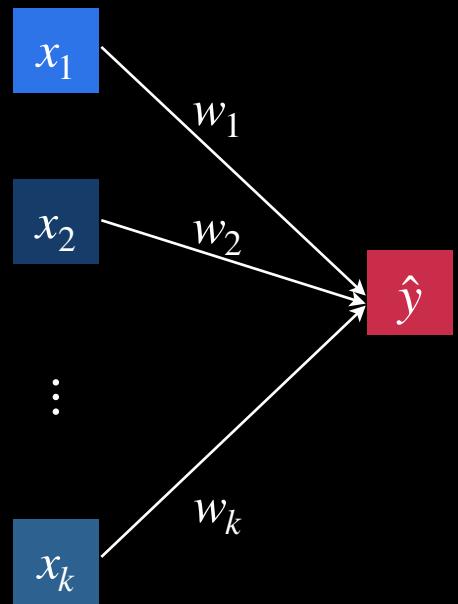
$$a = \sum_{i=1}^k w_i x_i \quad \dots \quad \text{Пре-активация}$$
$$\hat{y} = f(a) \quad \dots \quad \text{Активация}$$

↓
Функция активации

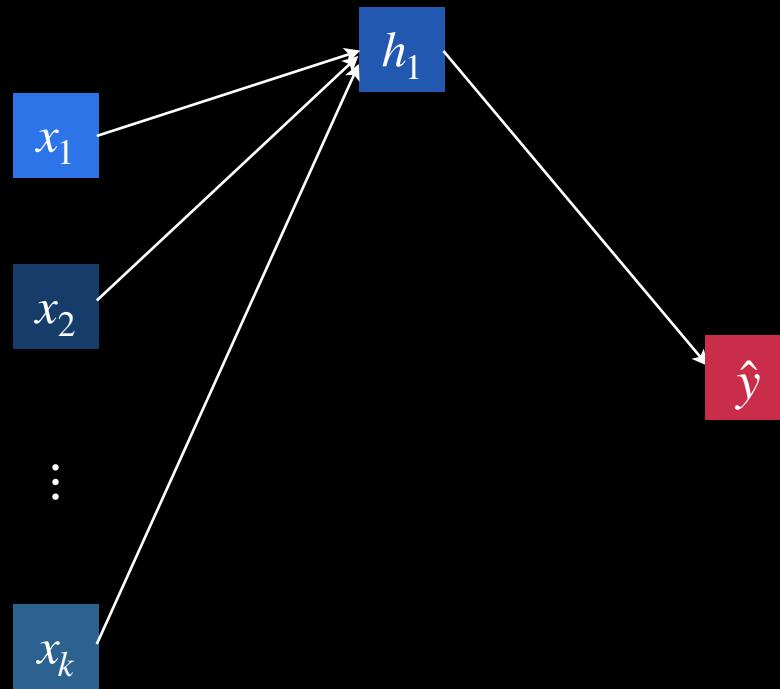
В матричной форме:

$$\hat{y} = f(\mathbf{w}^T \mathbf{x})$$

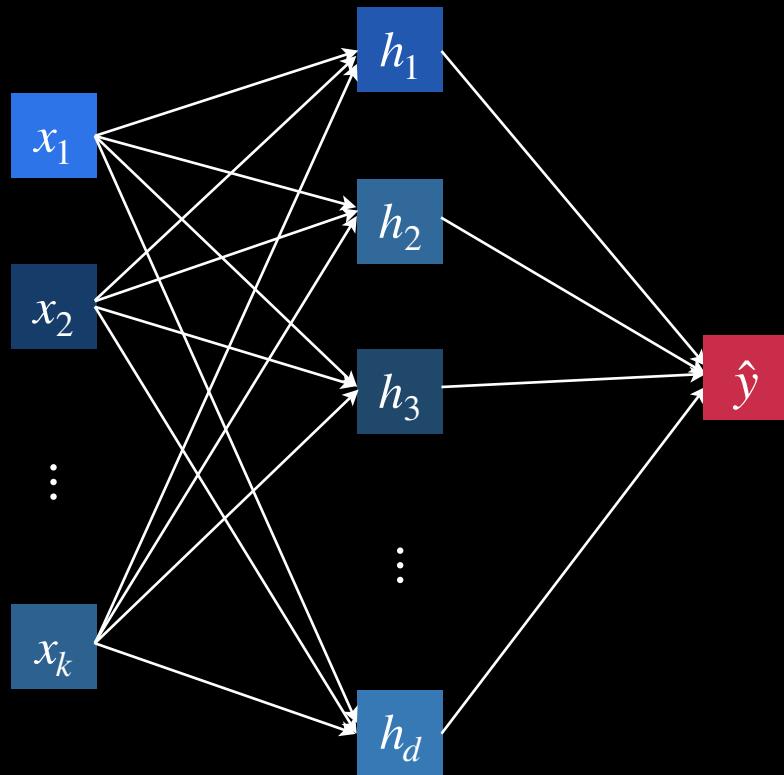
Однослойная нейронная сеть



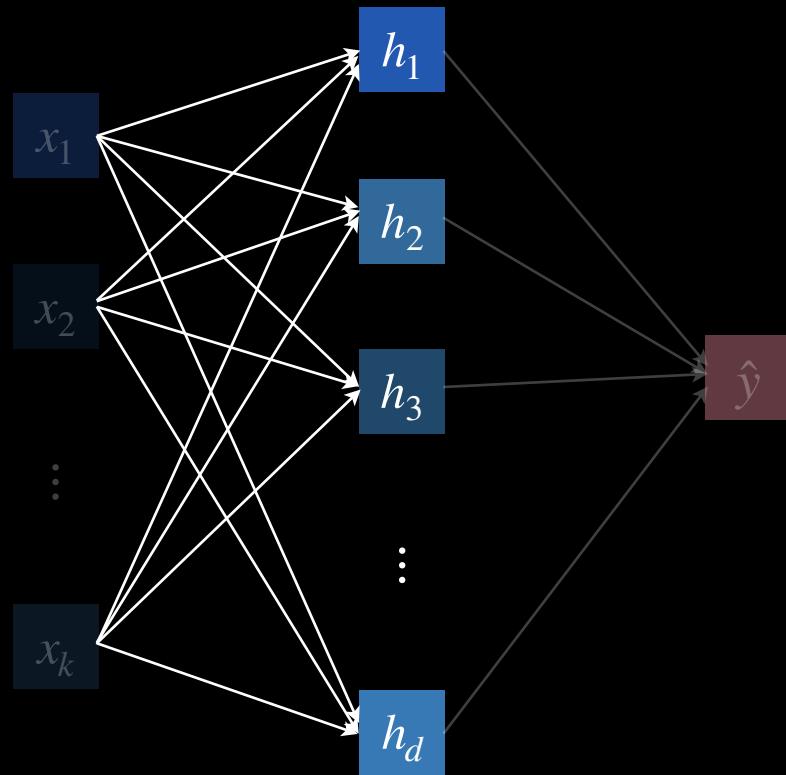
Добавление скрытого слоя



Добавление скрытого слоя

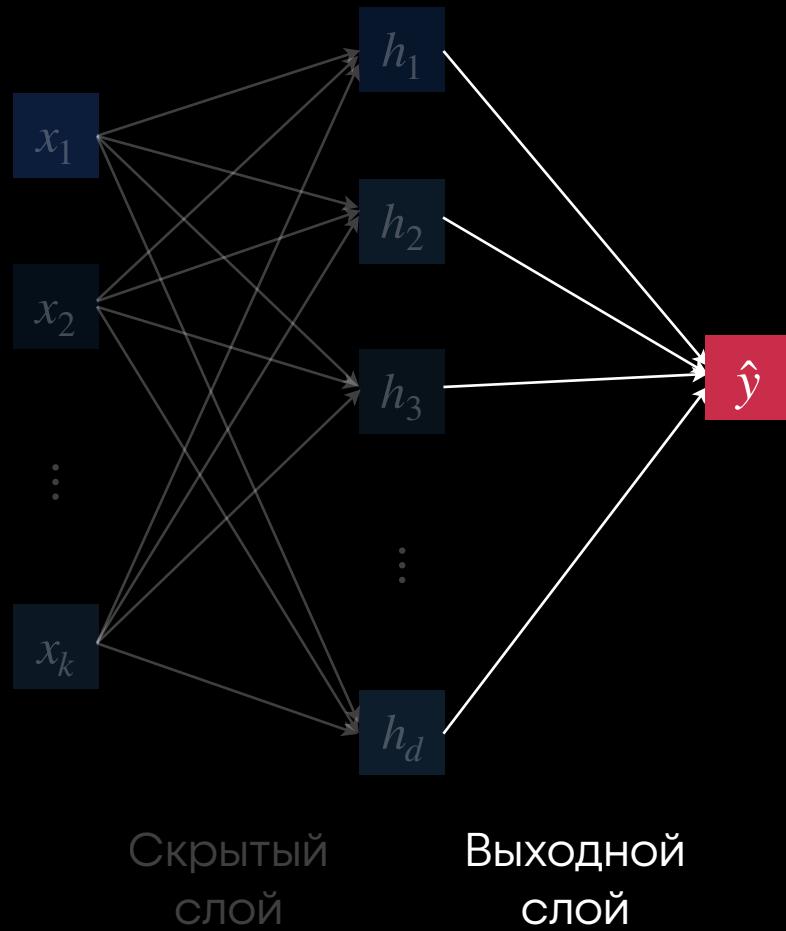


Добавление скрытого слоя

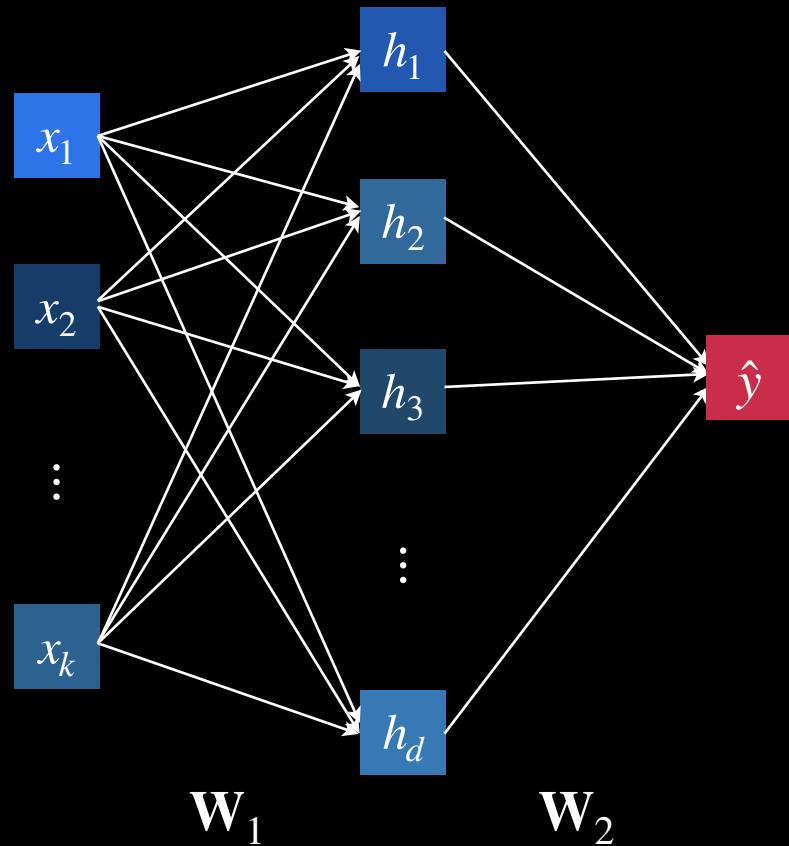


Скрытый
слой

Добавление скрытого слоя



Добавление скрытого слоя



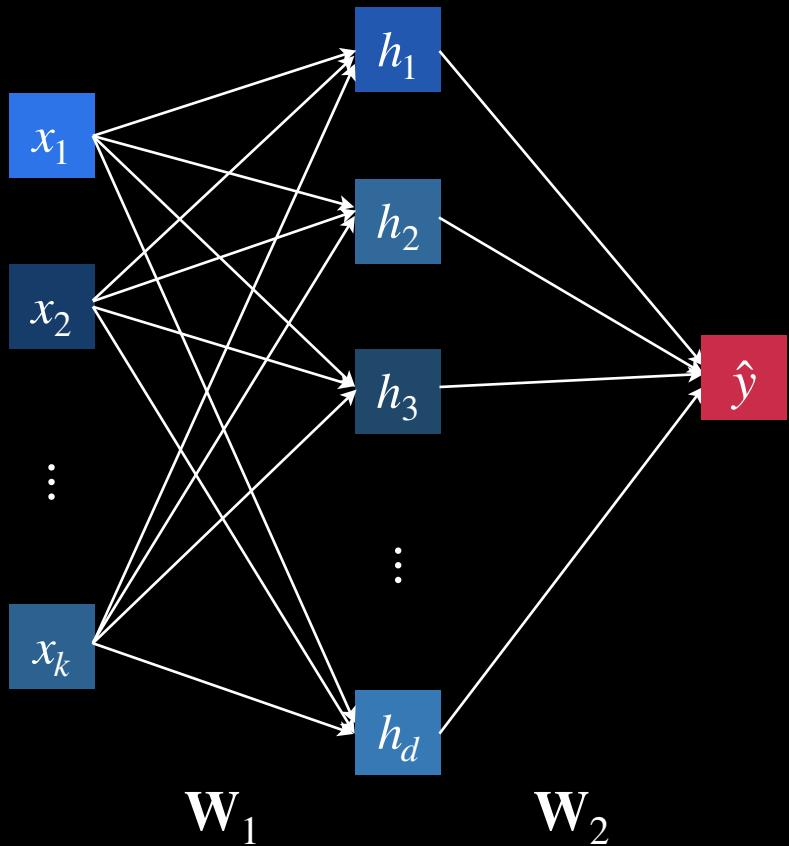
$$h_j = f_h \left(\sum_{i=1}^k w_{ji}^{(1)} x_i \right) \quad \text{Скрытые активации}$$
$$\hat{y}_j = f_o \left(\sum_{i=1}^d w_{ji}^{(2)} h_i \right) \quad \text{Выход сети}$$

В матричной форме:

$$\mathbf{h} = f_h (\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o (\mathbf{W}_2 \mathbf{h})$$

Добавление скрытого слоя



В матричной форме:

$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

Для некоторых нелинейных

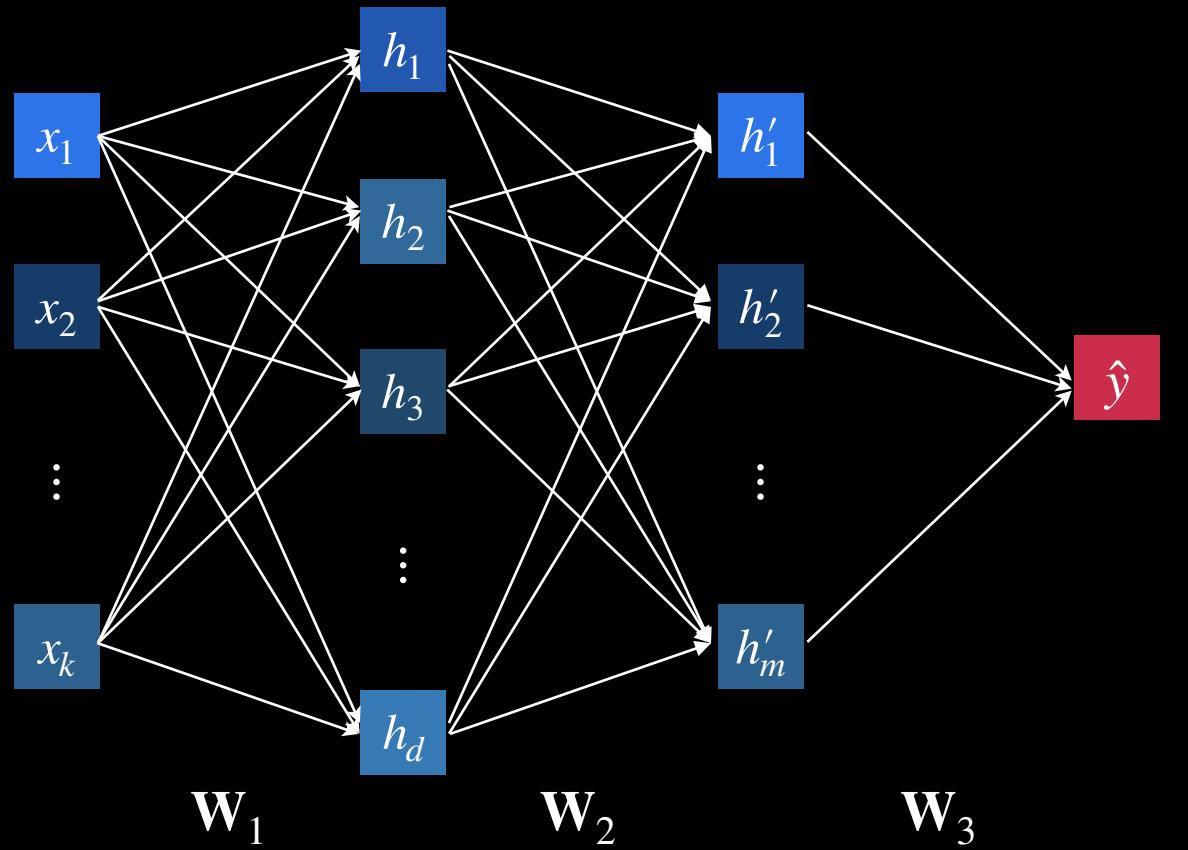
функций активации f_h

(например, для сигмоиды)

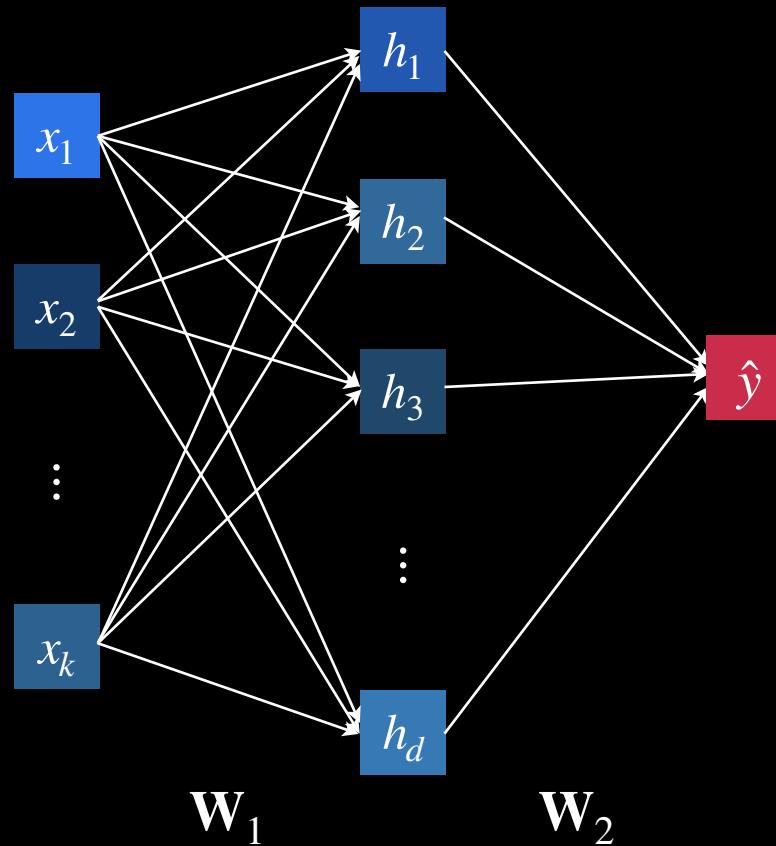
такая сеть —

универсальный аппроксиматор

Добавление скрытого слоя



Добавление скрытого слоя



В матричной форме:

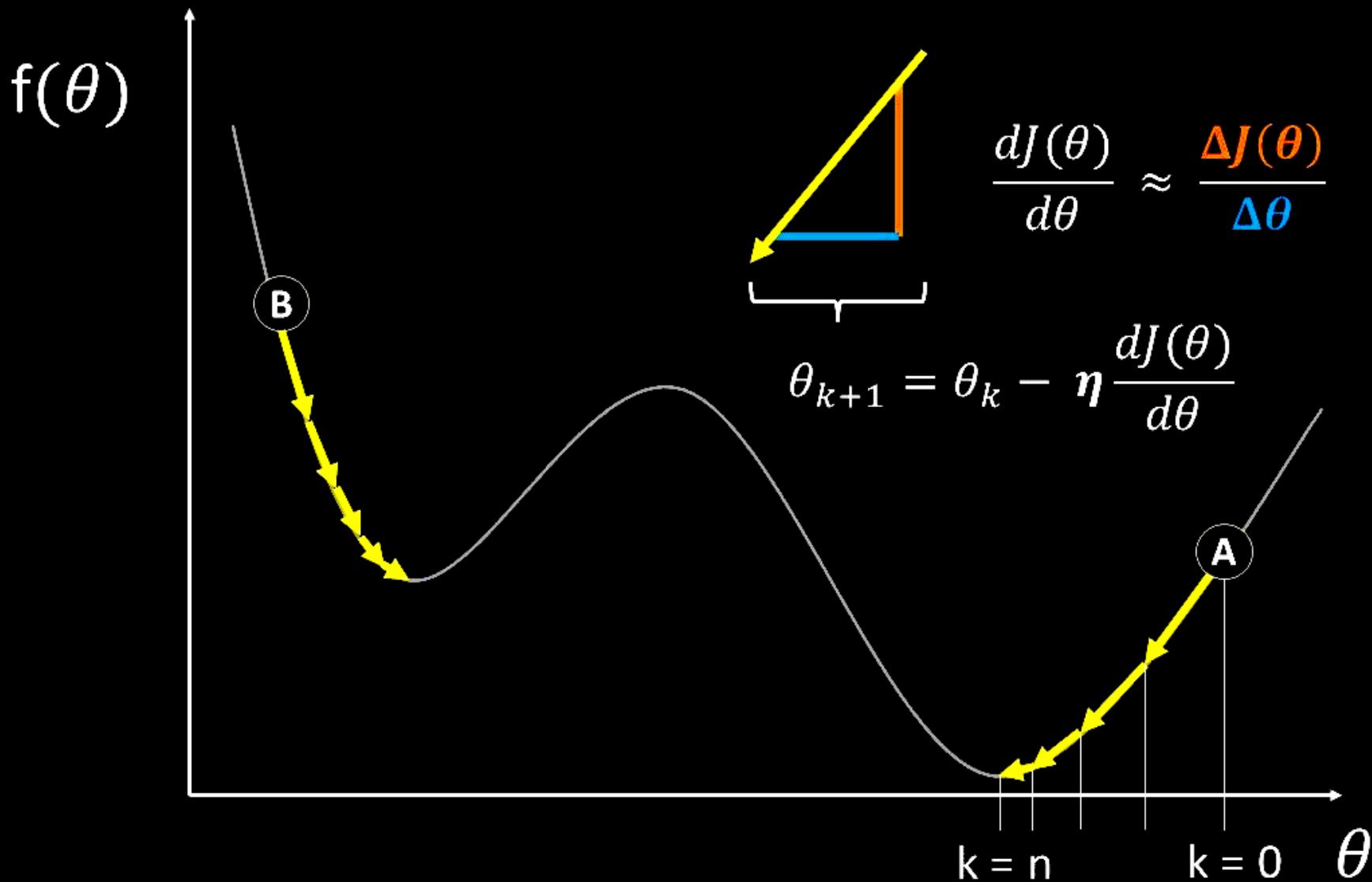
$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

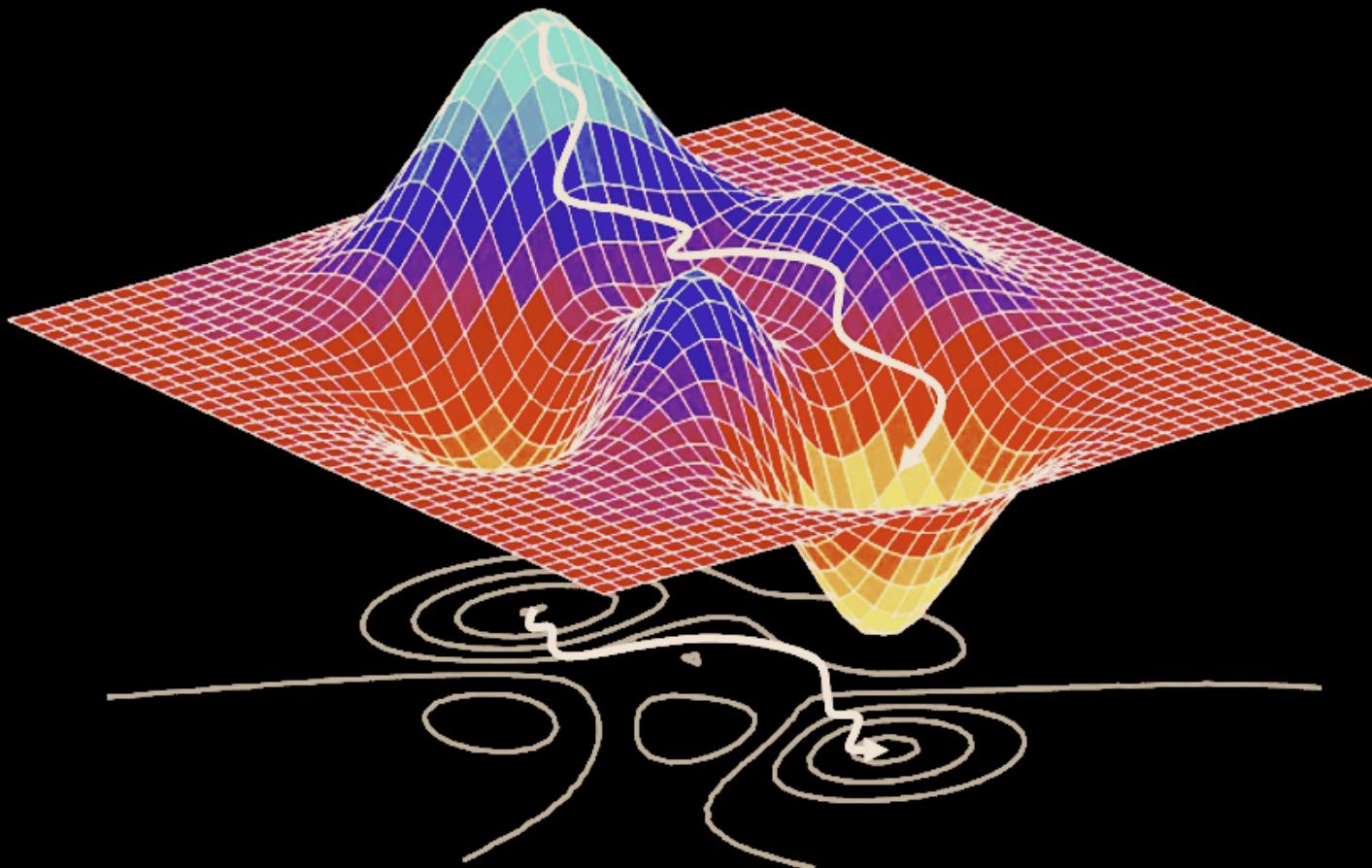
Как такое учить?

Производная и градиент

Производная и градиент



Производная и градиент



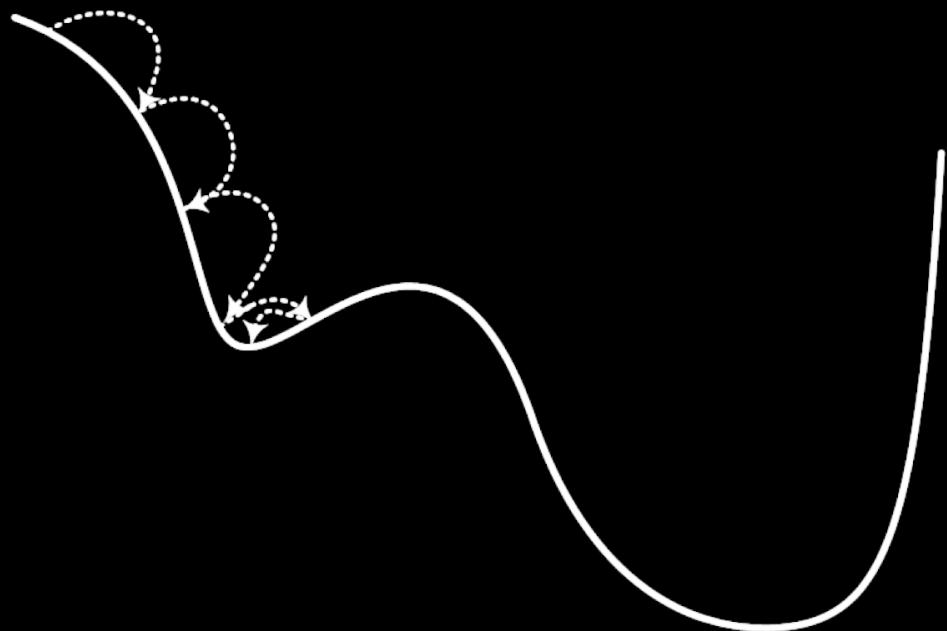
$E(\mathbf{W})$ — функция, которую хотим минимизировать

Градиентный спуск:

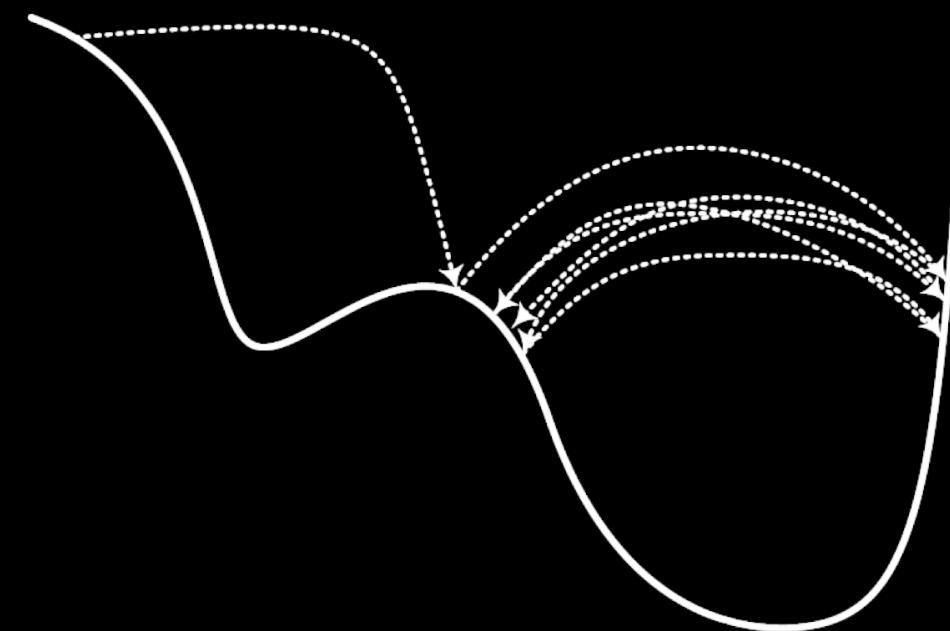
$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} - \eta \nabla E \quad \text{Градиент}$$

Размер шага

Выбор шага оптимизации

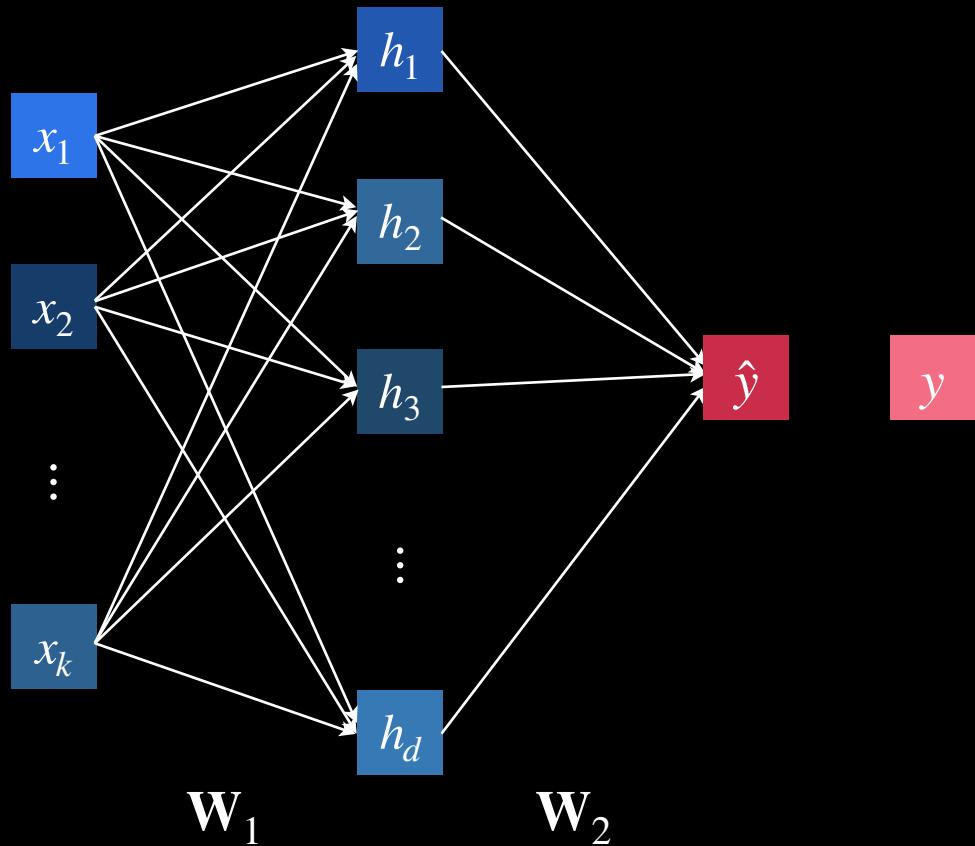


(а)



(б)

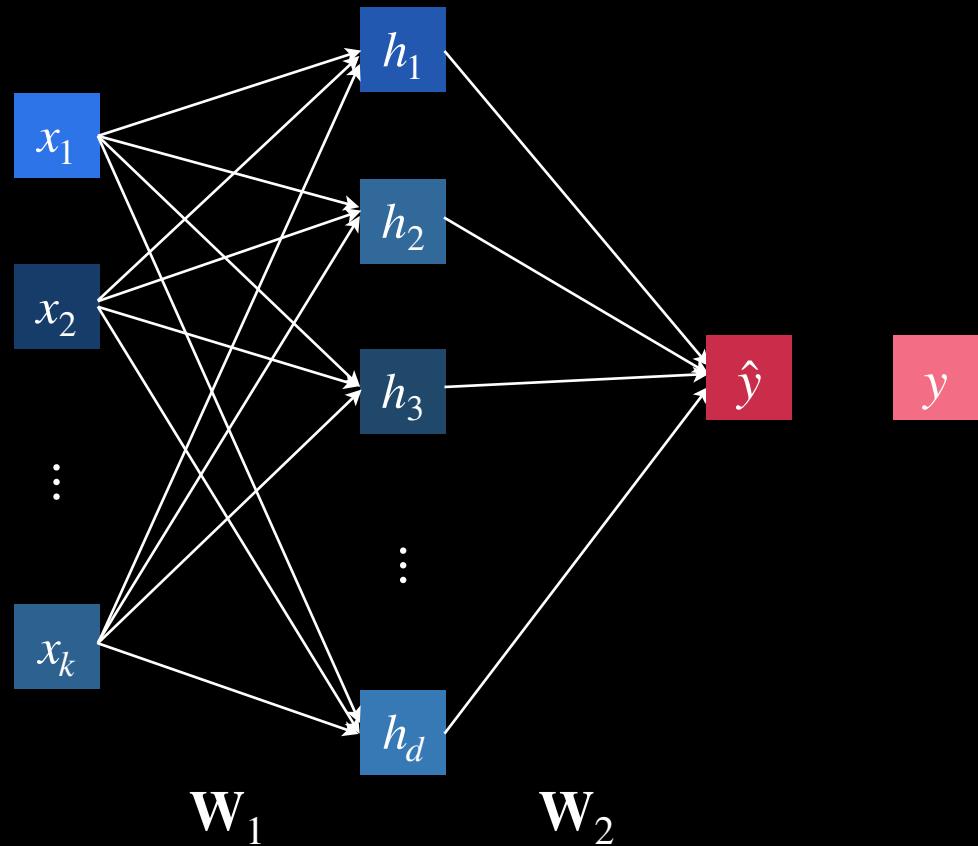
Обучение нейронных сетей



$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

Обучение нейронных сетей



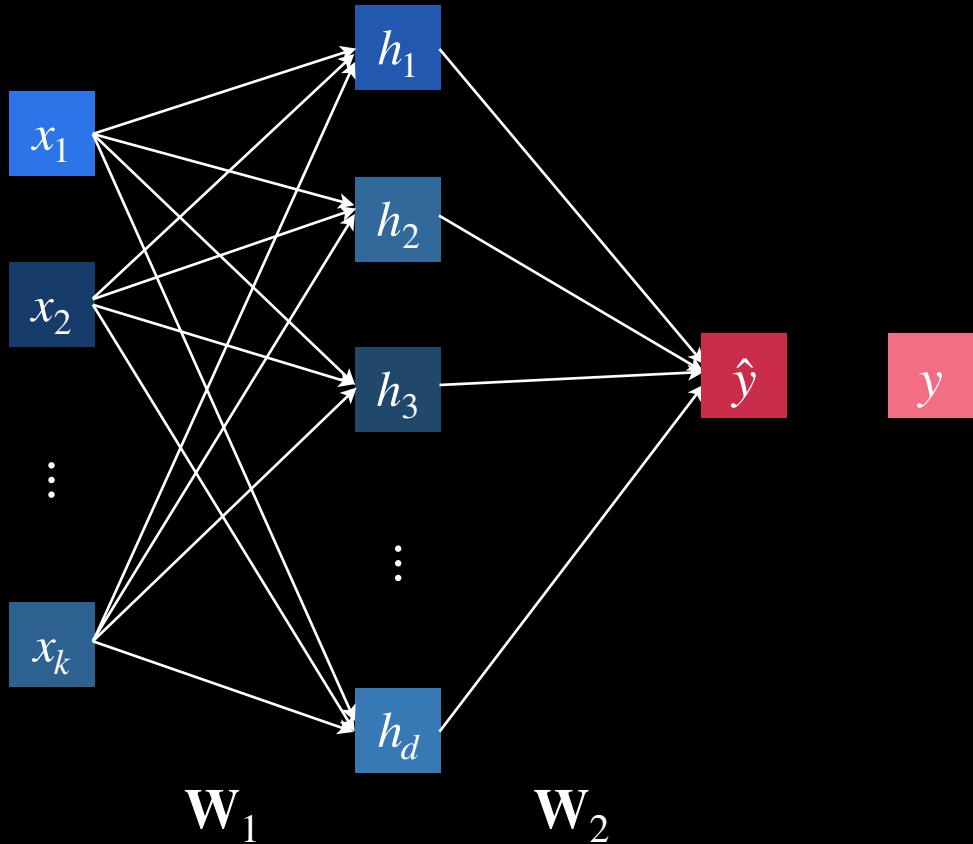
$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

$$E(\hat{y}, y)$$

Функция ошибки

Обучение нейронных сетей



$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

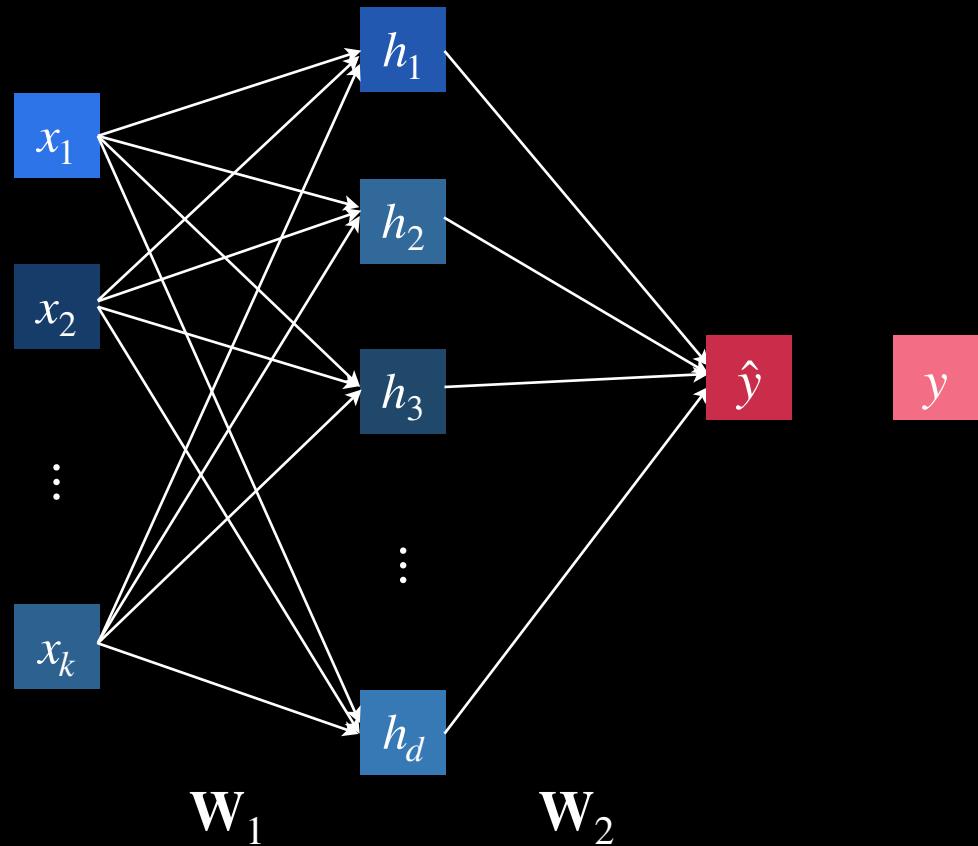
$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

$E(\hat{y}, y)$ Функция ошибки

Для задачи регрессии:

$$E(\hat{y}, y) = (\hat{y} - y)^2$$

Обучение нейронных сетей



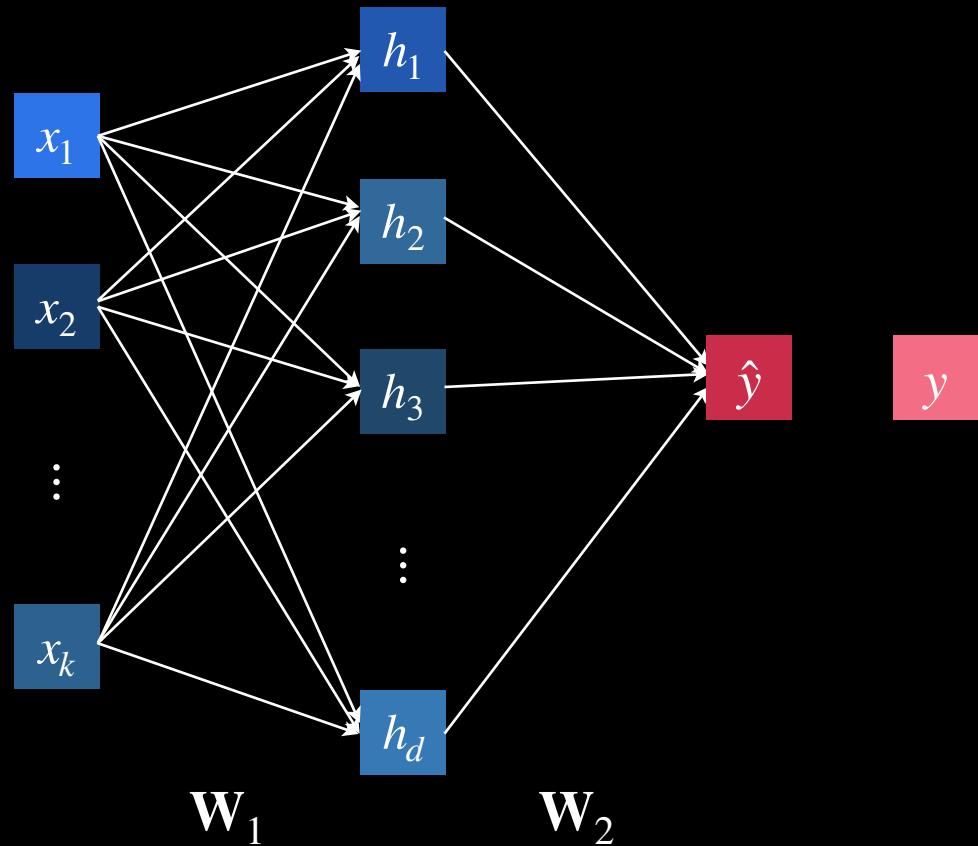
$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

$$E(\hat{y}, y)$$

Функция ошибки

Обучение нейронных сетей



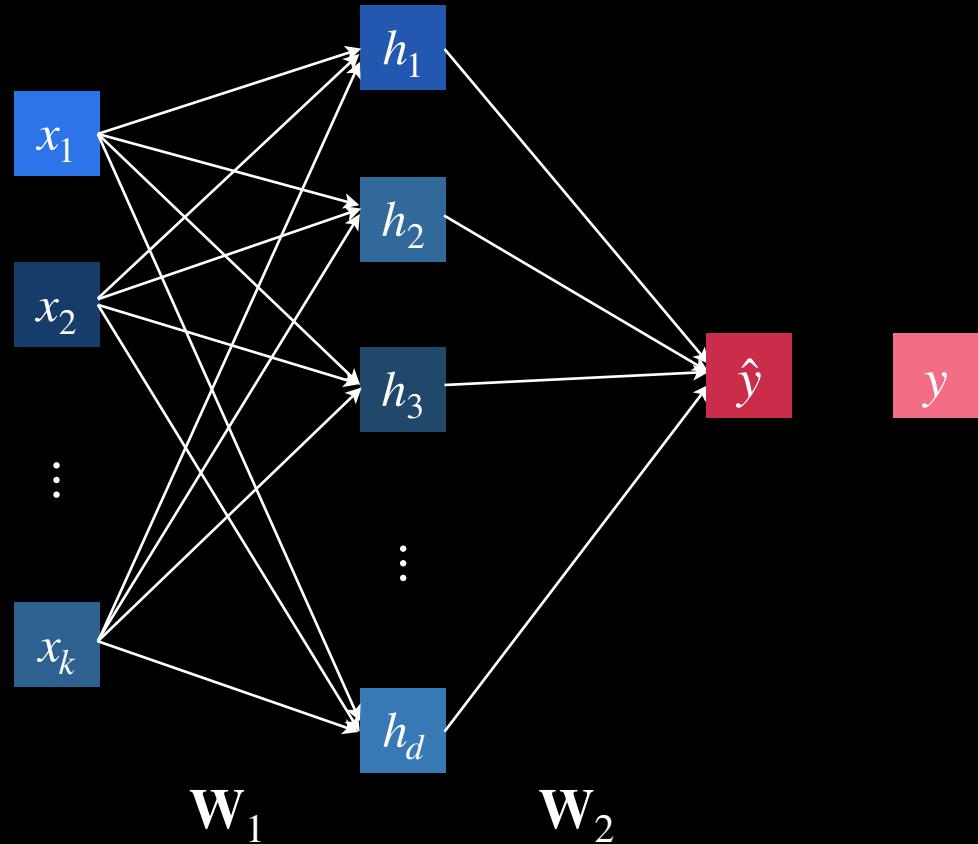
$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

$E(\hat{y}, y)$ Функция ошибки

$$E(\hat{y}, y) \longrightarrow \min_{\mathbf{W}}$$

Обучение нейронных сетей



$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

$E(\hat{y}, y)$ Функция ошибки

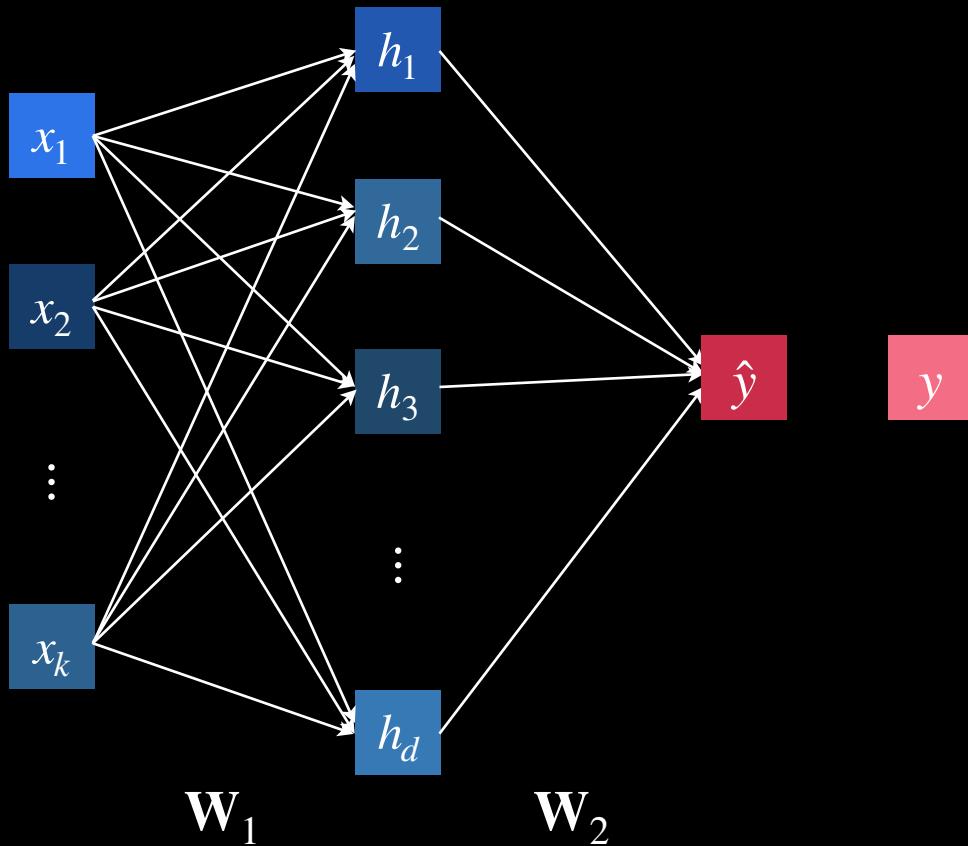
$$E(\hat{y}, y) \longrightarrow \min_{\mathbf{W}}$$

Трюк:

Функции активации — дифференцируемые

Функция ошибки — тоже

Обучение нейронных сетей



$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

$E(\hat{y}, y)$ Функция ошибки

$$E(\hat{y}, y) \longrightarrow \min_{\mathbf{W}}$$

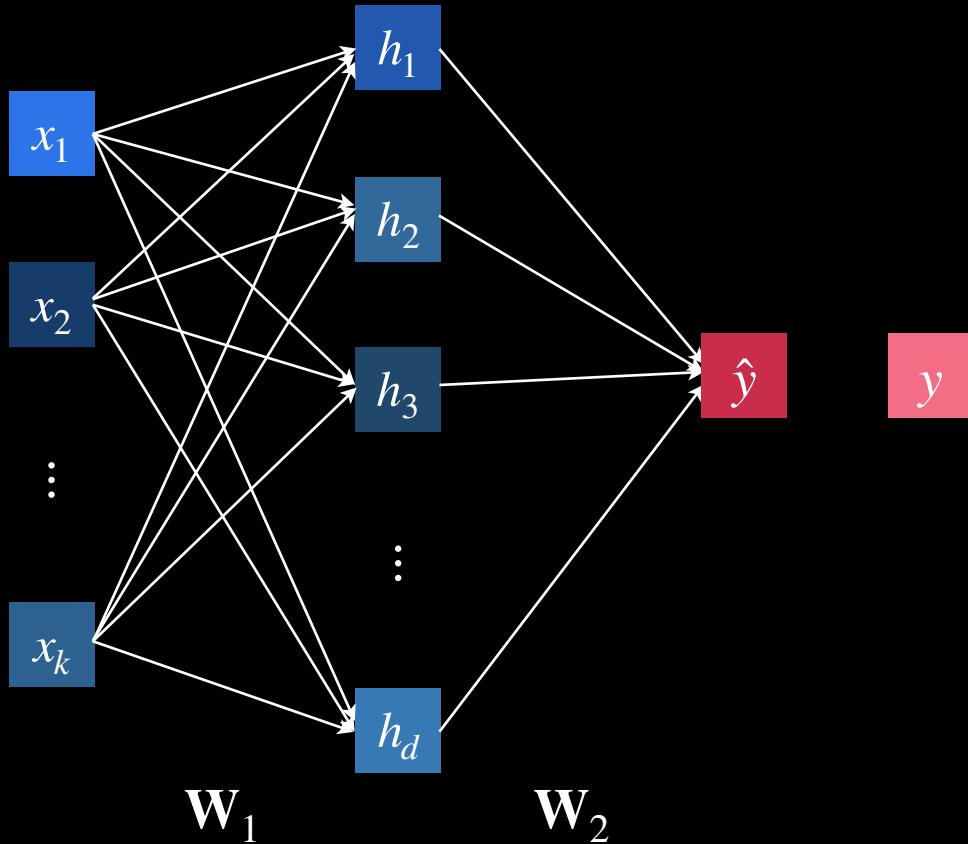
Трюк:

Функции активации — дифференцируемые

Функция ошибки — тоже

Значит можно посчитать производную ошибки по каждому из параметров!

Обучение нейронных сетей



$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

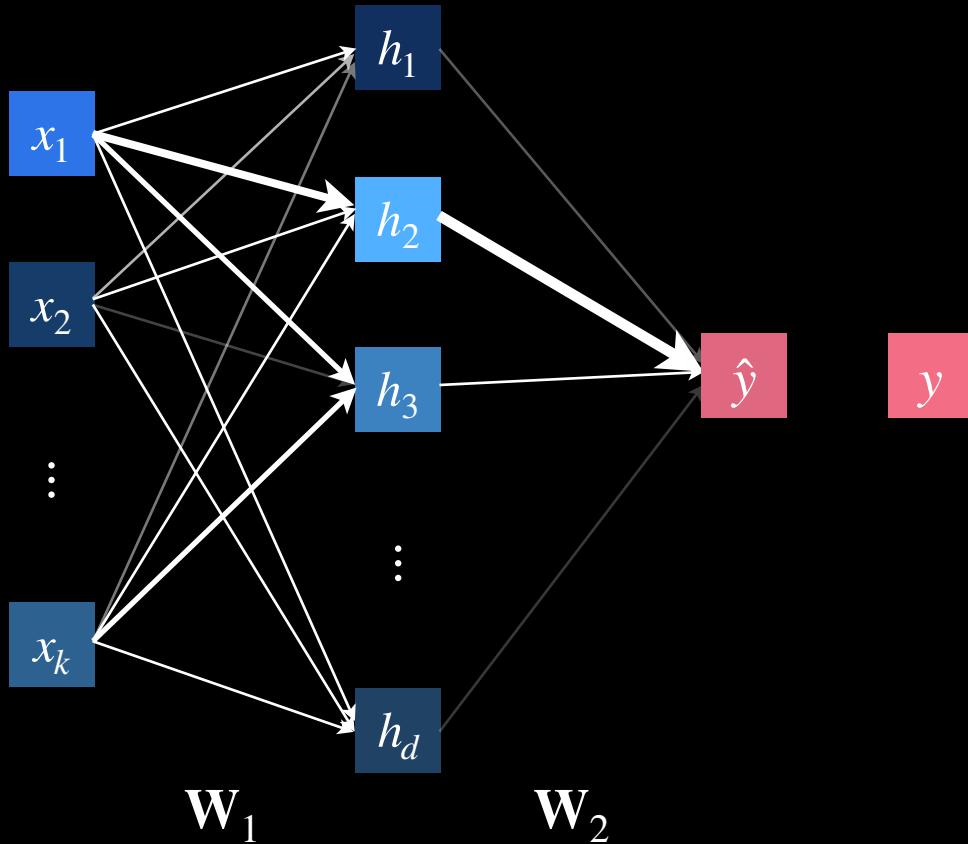
$E(\hat{y}, y)$ Функция ошибки

Градиентный спуск:

$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} - \eta \nabla E$ Градиент

..... Размер шага

Обучение нейронных сетей



$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

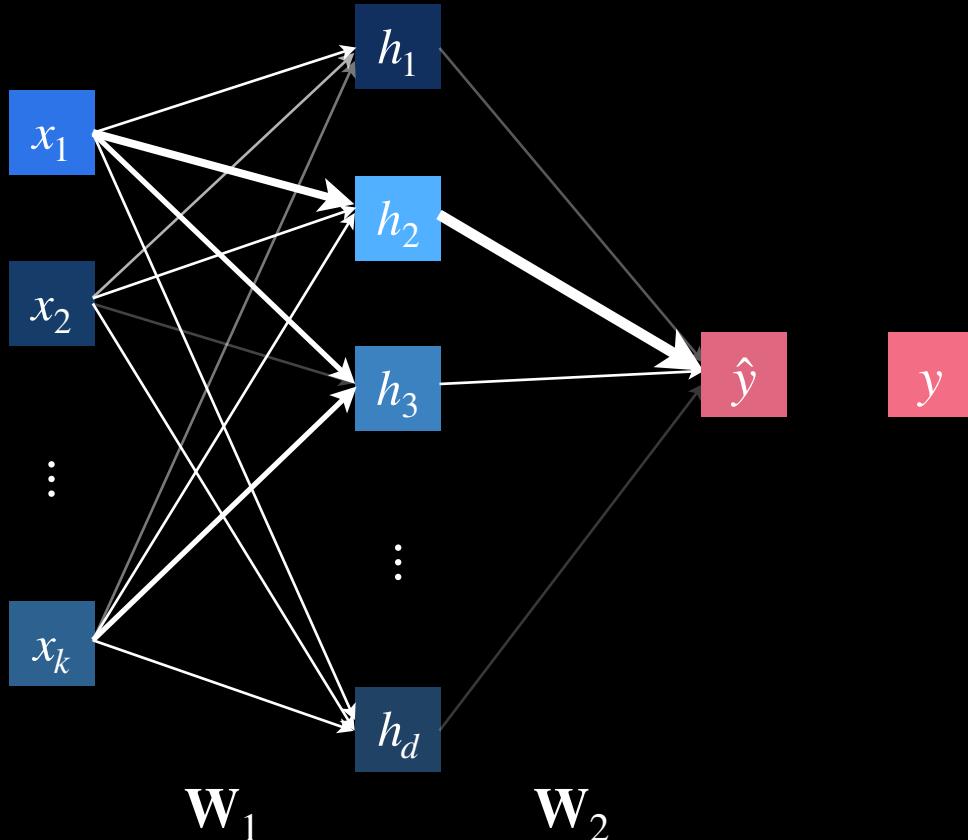
$$E(\hat{y}, y)$$

Градиентный спуск:

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} - \eta \nabla E$$

График градиента
Размер шага

Обучение нейронных сетей



$$\mathbf{h} = f_h(\mathbf{W}_1 \mathbf{x})$$

$$\hat{\mathbf{y}} = f_o(\mathbf{W}_2 \mathbf{h})$$

$$E(\hat{y}, y)$$

Функция ошибки

Градиентный спуск:

$$\mathbf{W}^{(t+1)} = \mathbf{W}^{(t)} - \eta \nabla E$$

Градиент

Размер шага