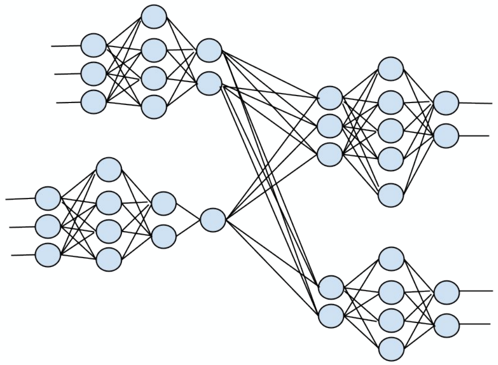
Введение в нейронные сети

Лектор — Дахова Елизавета

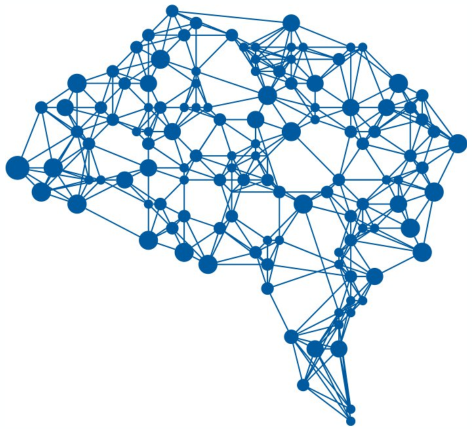
Что такое ~~эти ваши~~ нейронные сети?

Нейронная сеть — математическая модель,
а также ее программное или аппаратное воплощение,
построенная по принципу организации и функционирования
биологических нейронных сетей — сетей нервных клеток живого организма.

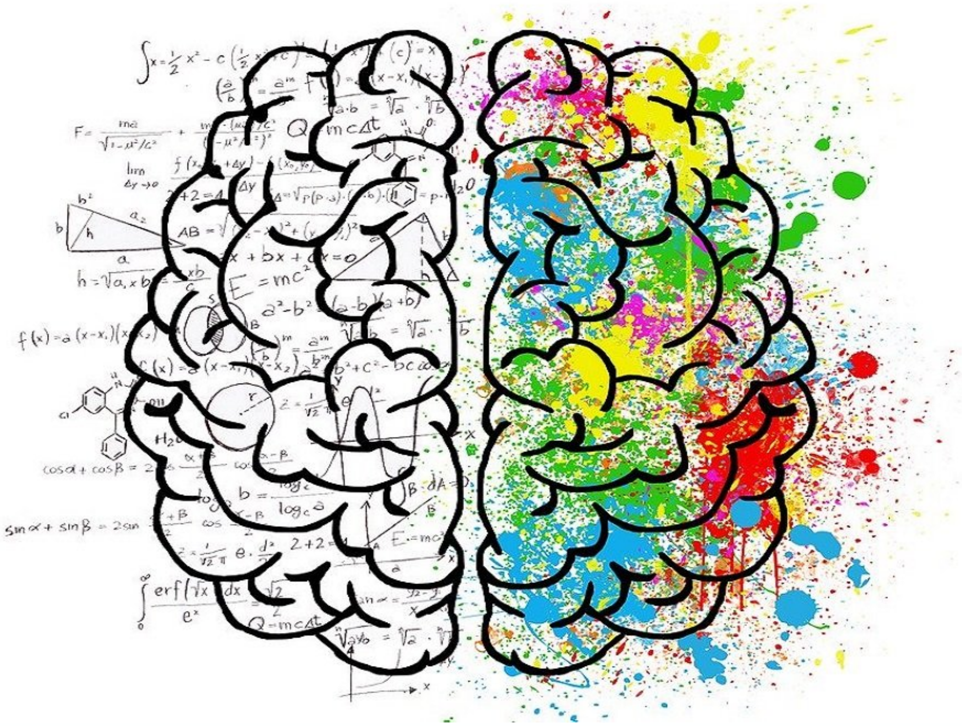
Это понятие возникло при изучении процессов, протекающих в мозге.



~



Мозг



Мозг – это

- Система обработки информации
- Сложный, нелинейный, параллельный компьютер

Решает множество сложных задач из области распознавания образов, обработки сигналов и прочее

Примеры

Задача распознавания знакомого лица в толпе займет 0.1-0.2 секунды

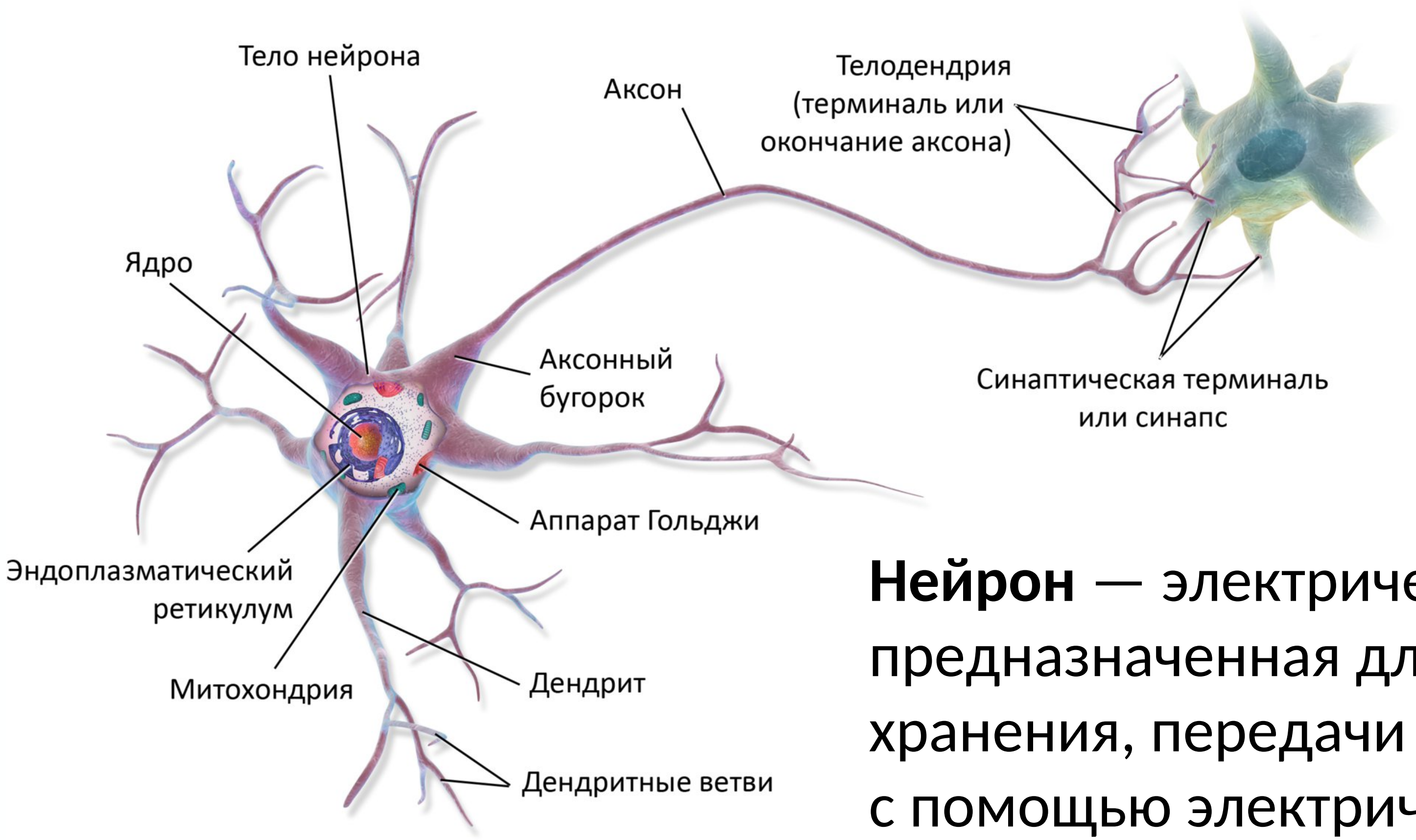
Примеры

- Эхо-локация:

летучая мышь получает информацию про положение, размер, скорость объекта и др..

Почему мы справляемся со столь сложными задачами?

Биологический нейрон



Структурно-функциональная единица нервной системы — это нейрон.

Нейрон — электрически возбудимая клетка, предназначенная для приёма, обработки, хранения, передачи и вывода информации с помощью электрических и химических сигналов.

Почему мы справляемся со столь сложными задачами?

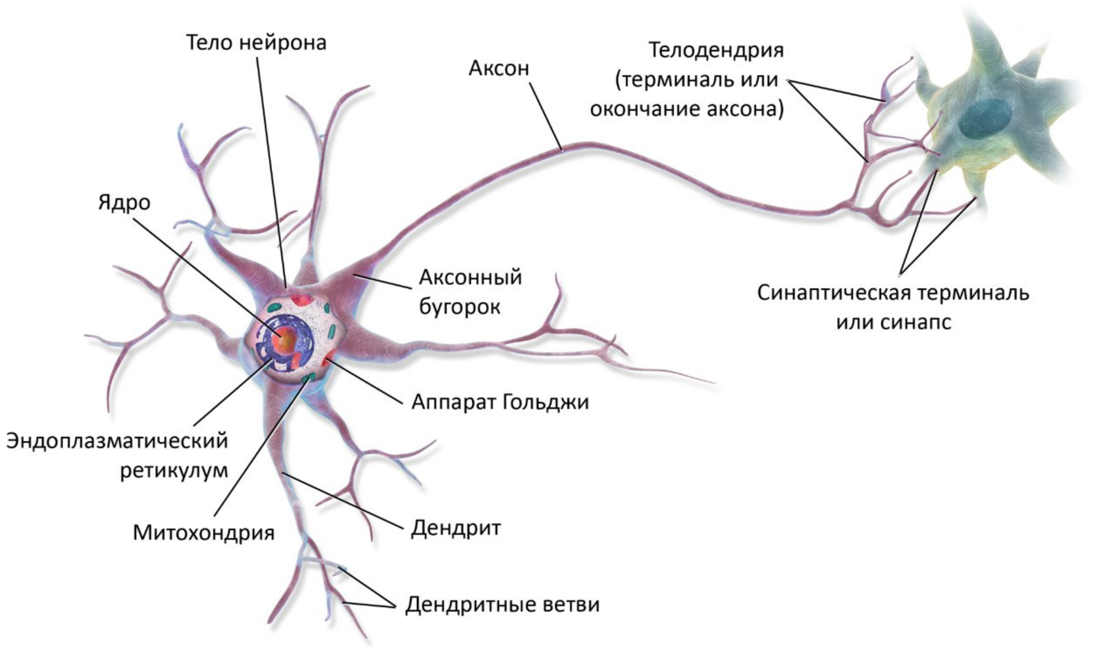
Опыт → Нейронная сеть → Способность решать сложные задачи

Обучение / training

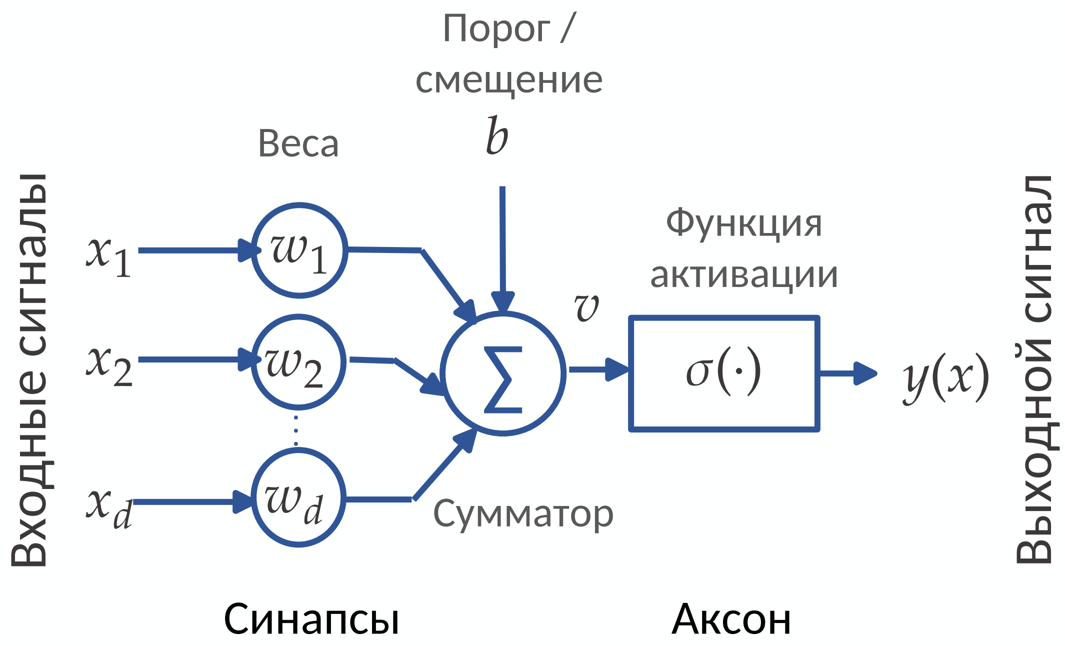
Применение / inference

Бионика: нейронные сети

Биологический нейрон



Искусственный нейрон



Историческая справка

1943 г. - Мак-Каллок и Питц

- Модель нейрона как «*всё или ничего*»
- Принципиальная возможность выполнить любые вычисления сетью нейронов

1949 г. – Хебб

- Открытие принципа формирования нейронных связей, взаимодействия нейронов

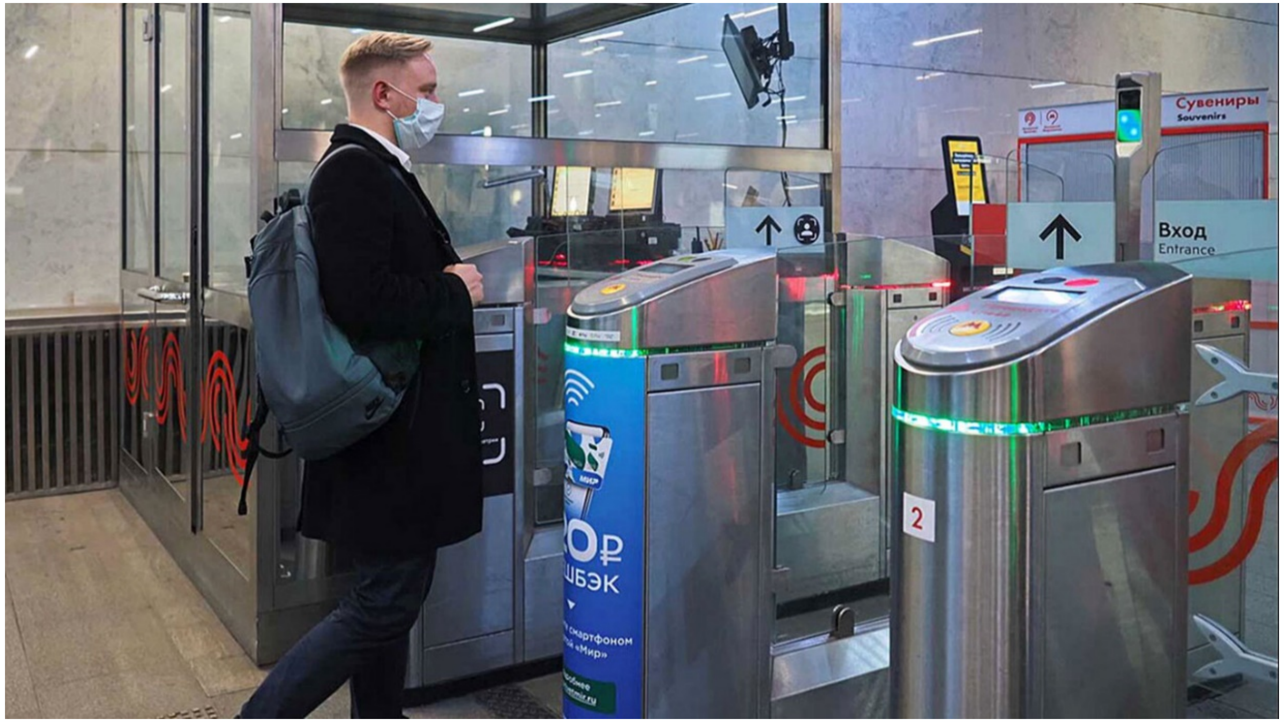
1958 г. – Розенблатт

- Открытие «персептрона» - устройства, моделирующего процесс человеческого восприятия

Области применения

Распознавание лиц / Face detection

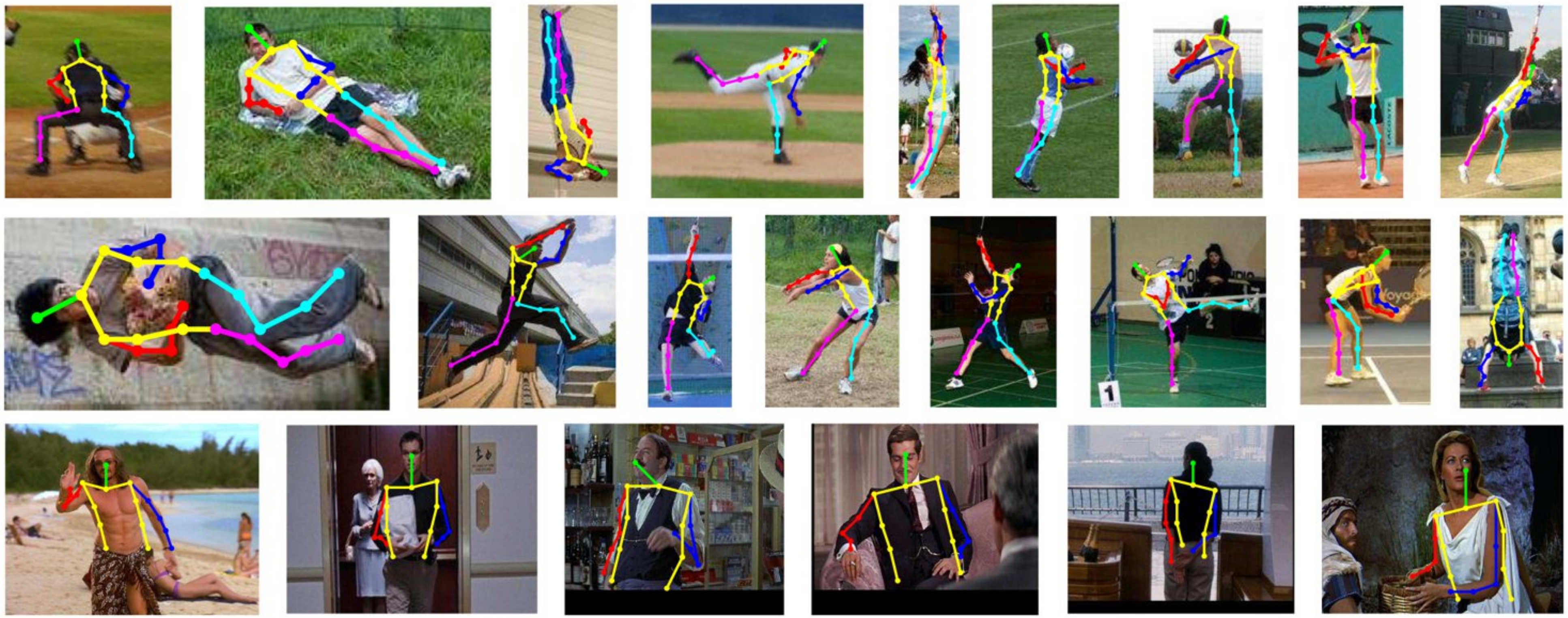
Цель — определить личность человека по снимку.



Области применения

Определение позы человека / Pose estimation

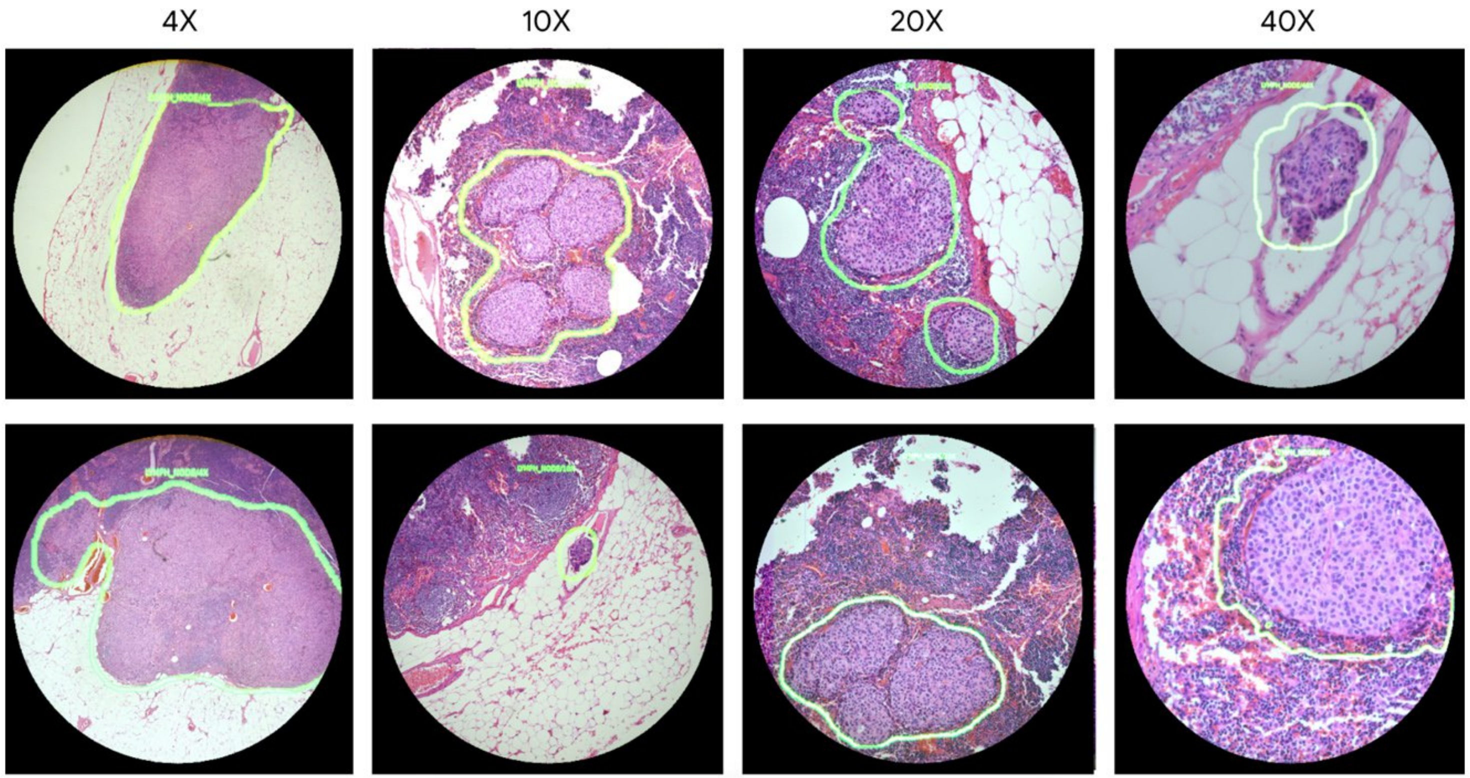
Цель — определить линии скелета человека .



Области применения

Сегментация опухоли / Tumor segementation

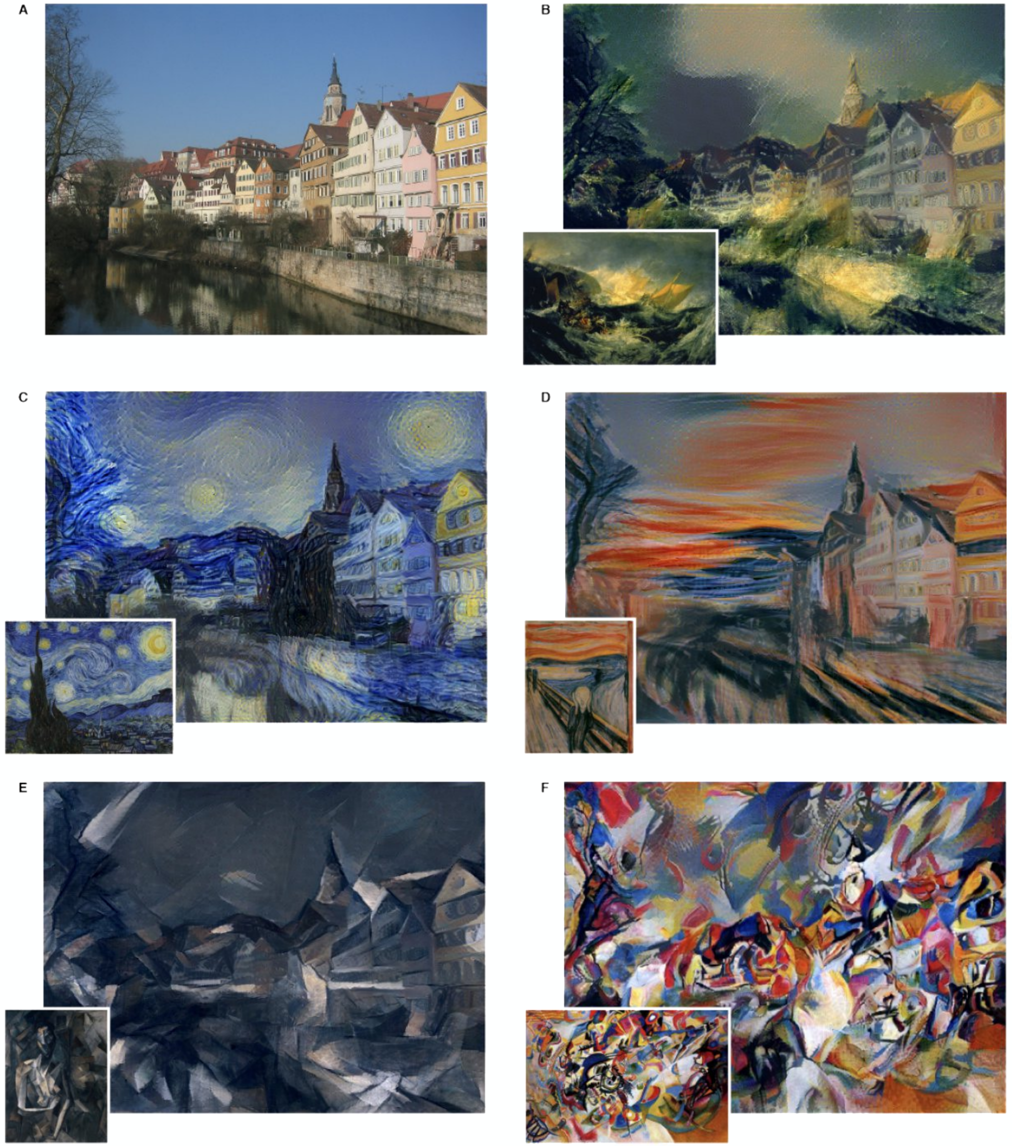
Цель — для каждого пикселя определить принадлежит ли он опухоли или нет.



Области применения

Перенос стиля / Style transfer

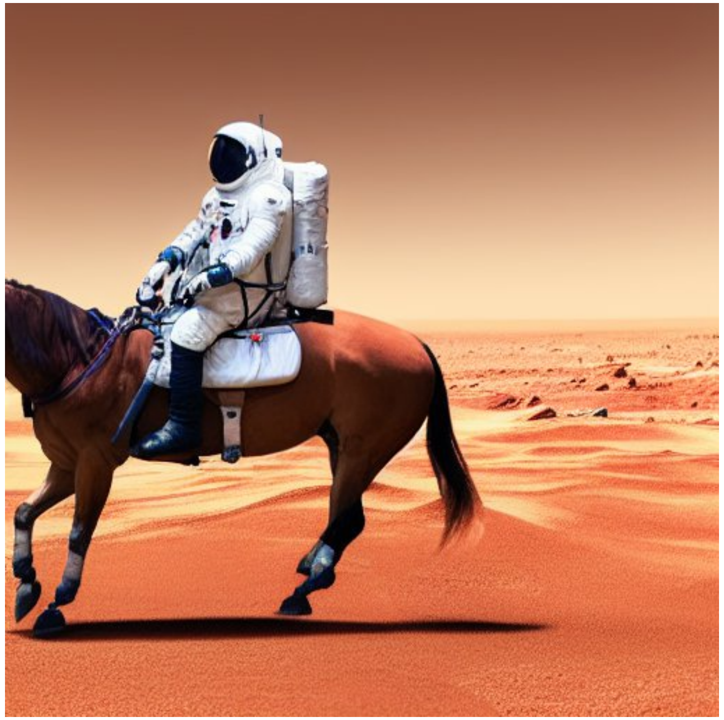
Цель — изменить изображение так,
чтобы новое изображение
сохранило структуру исходного,
но переняло стиль целевого изображения.



Области применения

Генерация изображений / Image generation

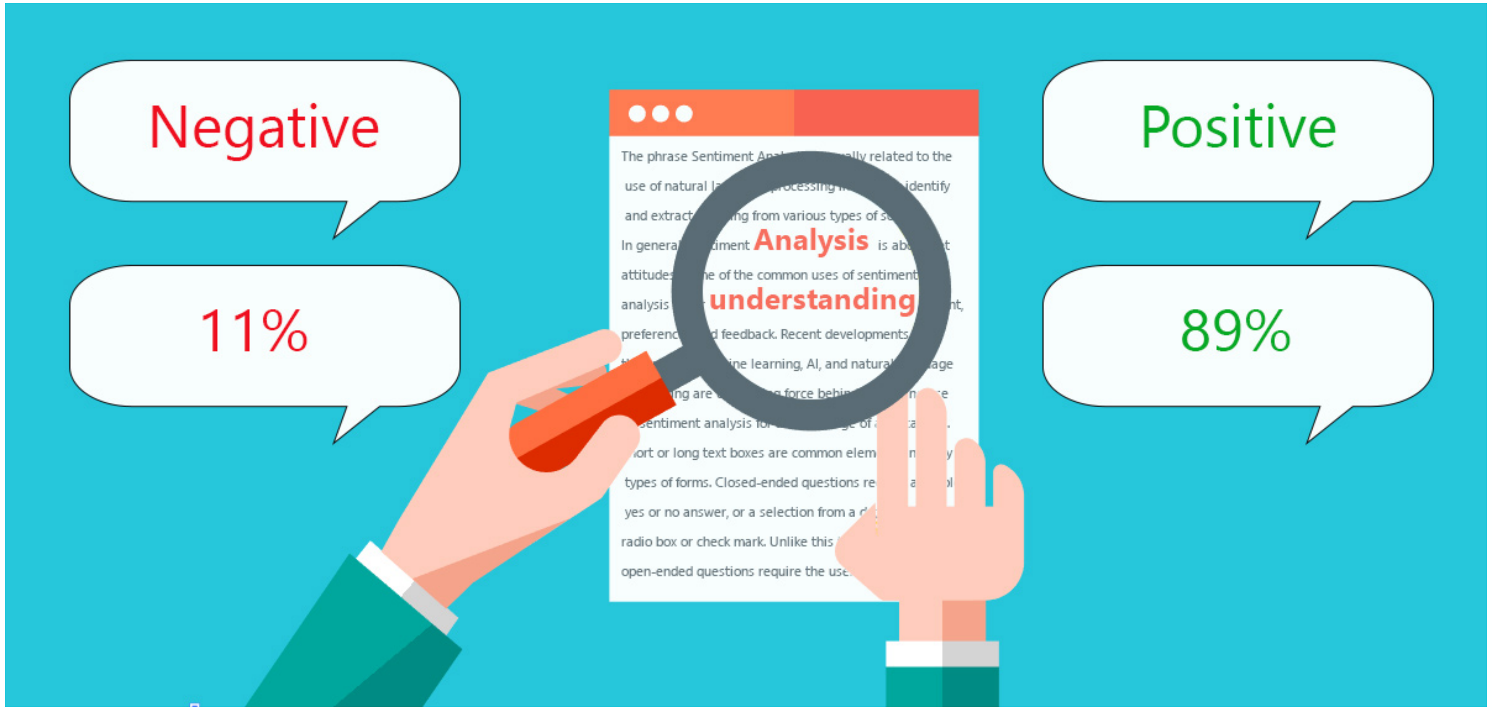
Цель — создание уникальных изображений.



Области применения

Определение тональности текста / Sentiment analysis

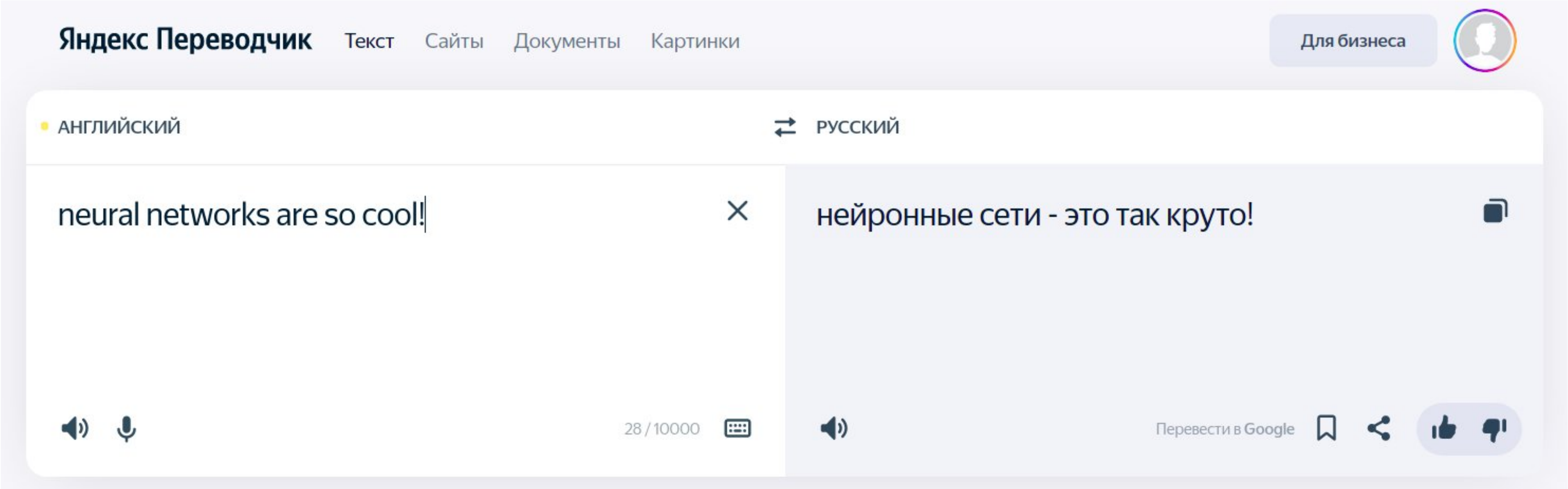
Цель — по определить является текст позитивным или негативным.



Области применения

Машинный перевод / Machine Language translation

Цель — перевести текст на одном языке на другой язык.



Области применения

Генерация текста / text generation

Цель — сгенерировать связный текст по соответствующей тематике.

Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{\acute{e}tale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

Lemma 0.2. *This is an integer $\mathcal{Z}$ is injective.*

Proof. See Spaces, Lemma ??. □

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset \mathcal{X}$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

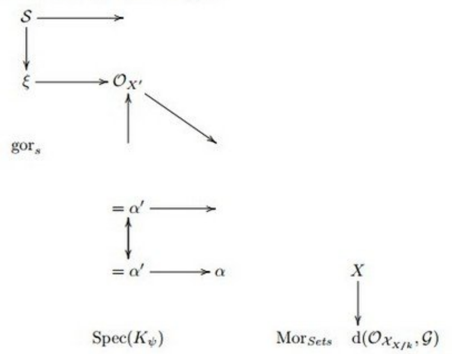
be a morphism of algebraic spaces over S and Y.

Proof. Let X be a nonzero scheme of X. Let X be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

(1) $\mathcal{F}$ is an algebraic space over S.
(2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □

This since $\mathcal{F} \in \mathcal{F}$ and $x \in \mathcal{G}$ the diagram



is a limit. Then $\mathcal{G}$ is a finite type and assume S is a flat and $\mathcal{F}$ and $\mathcal{G}$ is a finite type f_*. This is of finite type diagrams, and

- the composition of $\mathcal{G}$ is a regular sequence,
- $\mathcal{O}_{X'}$ is a sheaf of rings.

□

Proof. We have see that $X = \mathrm{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of X is an open neighbourhood of U. □

Proof. This is clear that $\mathcal{G}$ is a finite presentation, see Lemmas ??.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \quad \text{-1}(\mathcal{O}_{X_{\acute{e}tale}}) \longrightarrow \mathcal{O}_{X_\ell}^{-1}\mathcal{O}_{X_\lambda}(\mathcal{O}_{X_\eta}^{\overline{v}})$$

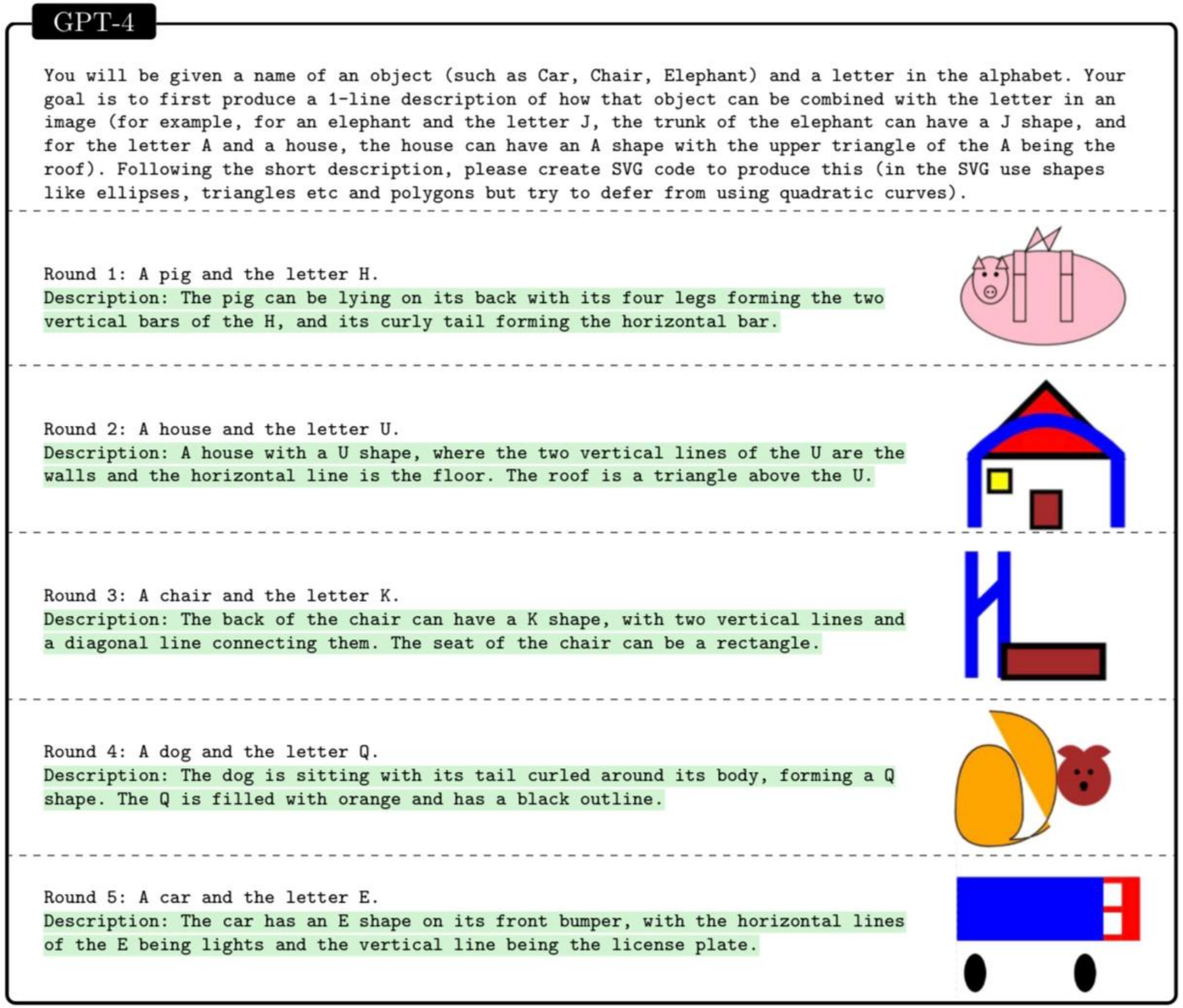
is an isomorphism of covering of $\mathcal{O}_{X_i}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that X is an isomorphism.
The property $\mathcal{F}$ is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma ??. This is a sequence of $\mathcal{F}$ is a similar morphism.

Области применения

Текстовые боты

/ text bots



Области применения

Распознавание речи / speech recognition

Цель — перевести человеческую речь в текст.

Области применения

Прогноз погоды / whether forecasting

Цель — предсказать температуру, осадки, скорость и направление ветра и тд.

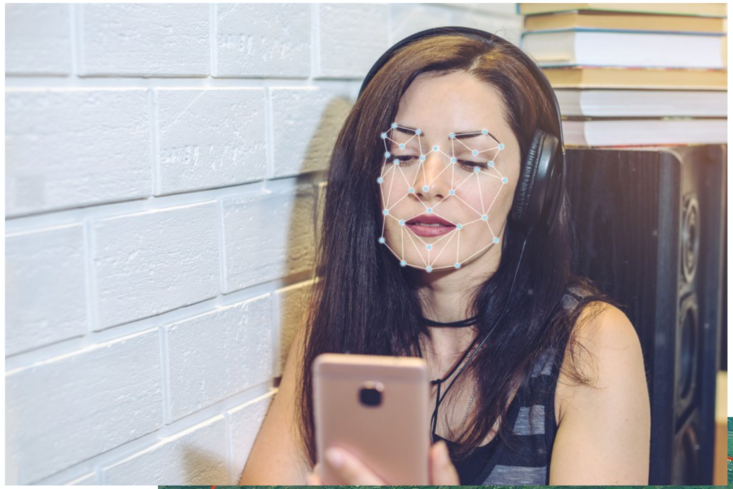
Области применения

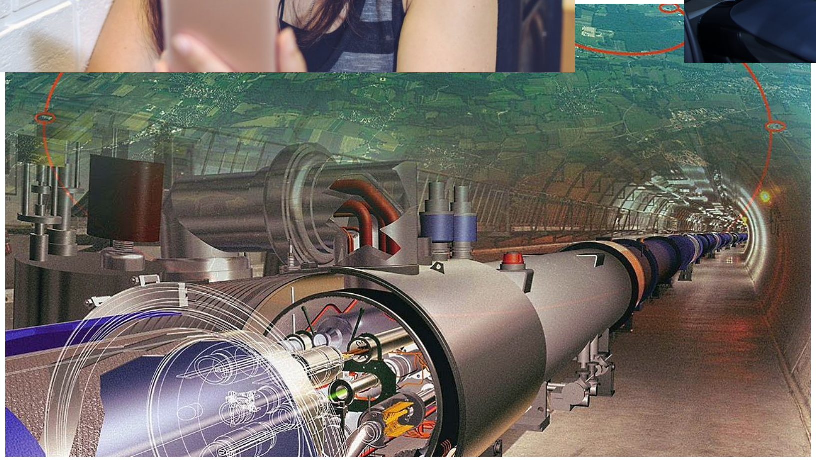
Предсказание продаж / sales prediction

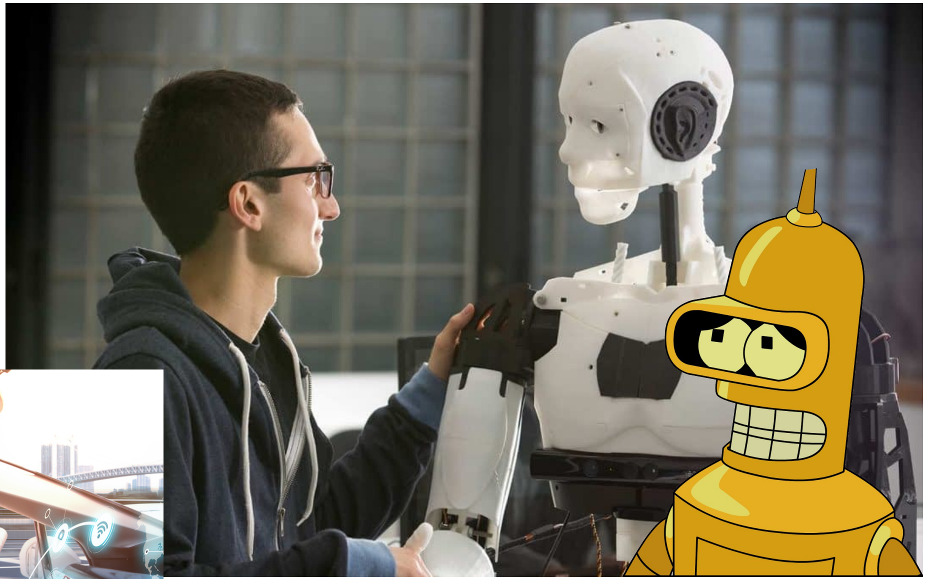
Цель — предсказать продажи товаров на некоторый период времени вперед.

Области применения

И многое другое…







Модель нейрона

Обозначим

$x \;=\; (x_1, \ldots, x_d)^T \in \mathbb{R}^d$ — один объект, где $x_1, \ldots, x_d$ — признаки;

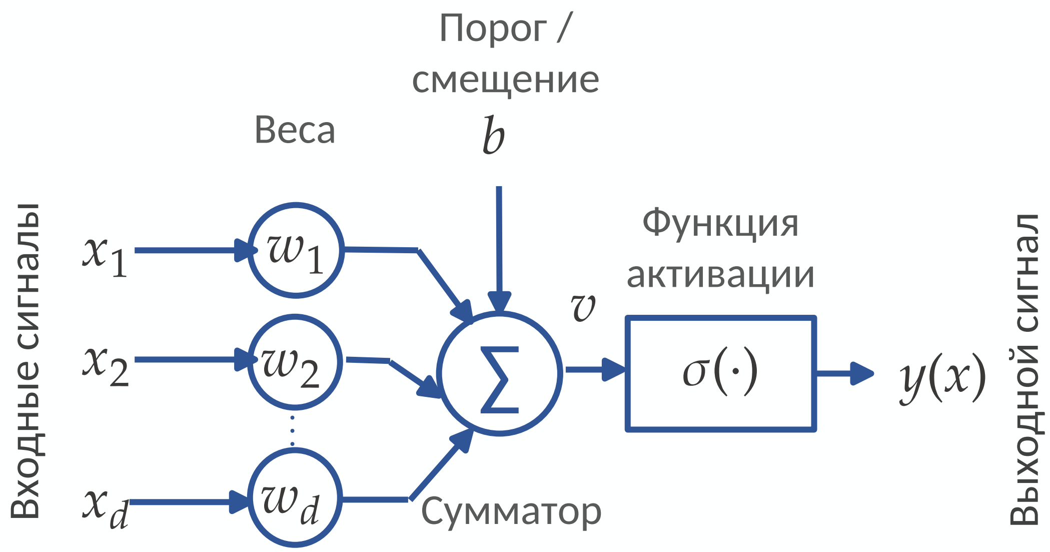
$w \;=\; (w_1, \ldots, w_d)^T \in \mathbb{R}^d$ — вектор весов;

$b \in \mathbb{R}$ — смещение.

Выход нейрона —

$$y(x) = \sigma(\langle x, w \rangle + b) = \sigma\left(\sum_{j=1}^{d} w_j x_j + b\right),$$

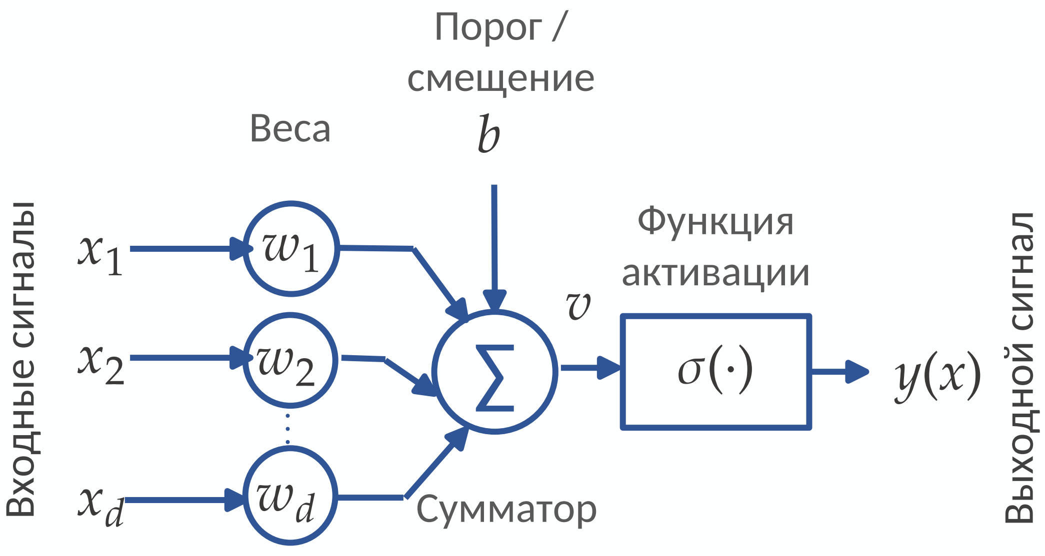
где σ — некоторая кус.-дифф. функция, назовем ее **функцией активации**.



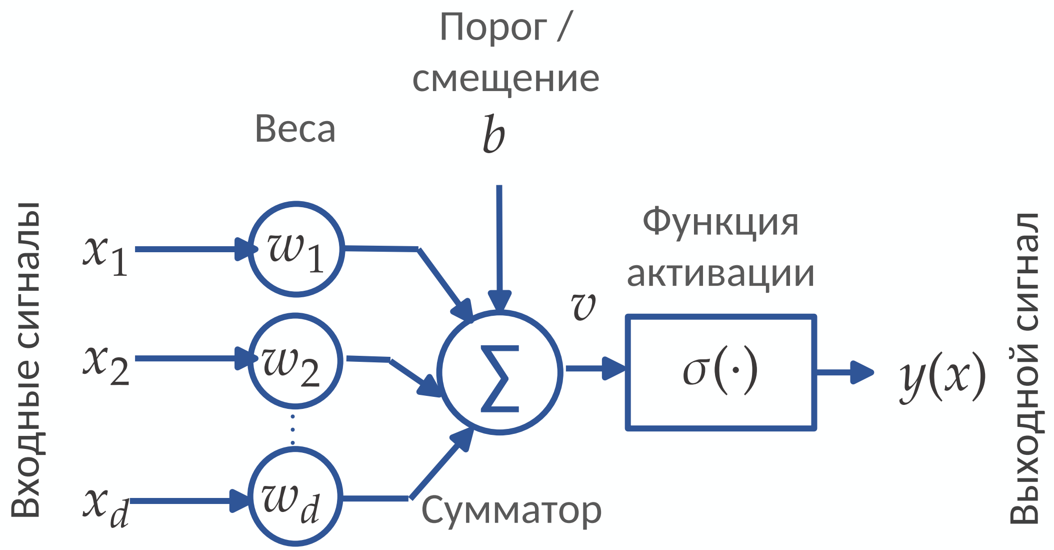
Что-то знакомое...

На что похожа эта формула?

$$y(x) = \sigma(\langle x, w \rangle + b) = \sigma\left(\sum_{j=1}^{d} w_j x_j + b\right)$$



Что-то знакомое...



На что похожа эта формула?

$$y(x) = \sigma(\langle x,w\rangle + b) = \sigma\left(\sum_{j=1}^{d} w_j x_j + b\right)$$

Линейная регрессия для 1 эл.

$$y(x) = \langle x,w\rangle + b = \sigma(\langle x,w\rangle + b) = \sigma\left(\sum_{j=1}^{d} x_j w_j + b\right),$$

где $\sigma(z) = z$ — линейная ф-я.

Логистическая регрессия для 1 эл. (будет на 3 курсе)

$y(x) = \sigma(\langle x,w\rangle + b) = \sigma\left(\sum_{j=1}^{d} x_j w_j + b\right),$ где $\sigma(z) = \frac{1}{1+e^{-z}}$ — логистическая сигмоида.

Пуассоновская регрессия для 1 эл. (будет на 3 курсе)

$y(x) = \sigma(\langle x,w\rangle + b) = \sigma\left(\sum_{j=1}^{d} x_j w_j + b\right),$ где $\sigma(z) = e^z$.

Решение задачи регрессии

Линейная регрессия: $\hat{y} = Xw + b$, т.е. $\hat{y}_i = \sum_{j=1}^{d} x_{ij} w_j + b$

где $X = (x_{ij})_{ij}$ — матрица входных данных, $i \in \{1, \ldots, n\}, j \in \{1, \ldots, d\}$

$\hat{y} = (\hat{y}_1, \ldots, \hat{y}_d)^T$ — вектор предсказания,

$w = (w_1, \ldots, w_d)^T$ — вектор весов, b — сдвиг.

Задачу можно решить аналитически. А можно с помощью **градиентного спуска**.

Зададим функцию, которую мы хотим минимизировать

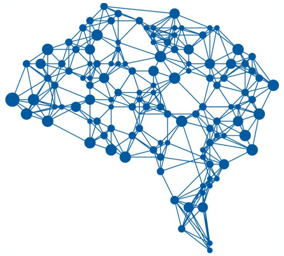
$$L = \frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 \text{ — MSE (Mean Squared Error)}$$

Пусть $\theta = (w, b)$, тогда оптимизация будет следующей:

$$\theta_t = \theta_{t-1} - \eta \nabla L(\theta_t), \quad \text{где } \eta \text{ — скорость обучения}$$

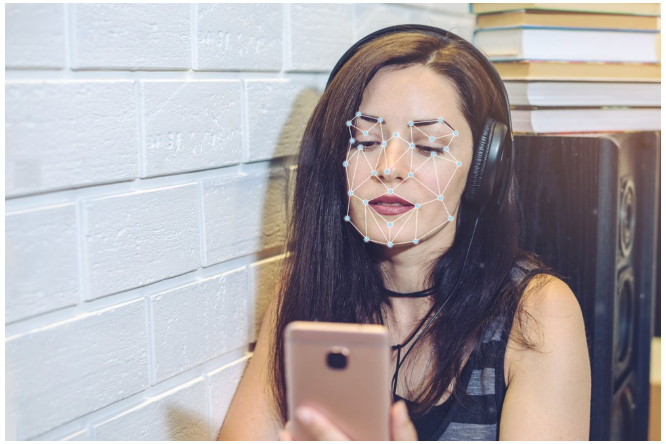
Более сложные задачи

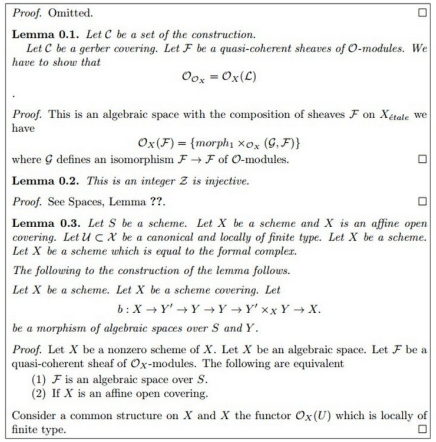
- У линейной и логистической регрессий ограниченная область применения.
- Для того, чтобы решить нелинейную задачу,
  нужно делать сложные преобразования с признаками.
- Один нейрон не справится со сложными задачами… ☹



Вспомним, что в нервной системе очень много нейронов.

Значит, будем использовать **больше нейронов**!





Proof. Omitted. □

Lemma 0.1. *Let $\mathcal{C}$ be a set of the construction.*

Let $\mathcal{C}$ be a gerber covering. Let $\mathcal{F}$ be a quasi-coherent sheaves of $\mathcal{O}$-modules. We have to show that

$$\mathcal{O}_{\mathcal{O}_X} = \mathcal{O}_X(\mathcal{L})$$

.

Proof. This is an algebraic space with the composition of sheaves $\mathcal{F}$ on $X_{étale}$ we have

$$\mathcal{O}_X(\mathcal{F}) = \{morph_1 \times_{\mathcal{O}_X} (\mathcal{G}, \mathcal{F})\}$$

where $\mathcal{G}$ defines an isomorphism $\mathcal{F} \to \mathcal{F}$ of $\mathcal{O}$-modules. □

Lemma 0.2. *This is an integer $\mathcal{Z}$ is injective.*

Proof. See Spaces, Lemma ??. □

Lemma 0.3. *Let S be a scheme. Let X be a scheme and X is an affine open covering. Let $\mathcal{U} \subset \mathcal{X}$ be a canonical and locally of finite type. Let X be a scheme. Let X be a scheme which is equal to the formal complex.*

The following to the construction of the lemma follows.

Let X be a scheme. Let X be a scheme covering. Let

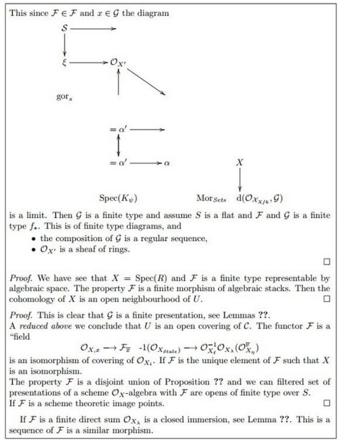
$$b : X \to Y' \to Y \to Y \to Y' \times_X Y \to X.$$

be a morphism of algebraic spaces over S and Y.

Proof. Let X be a nonzero scheme of X. Let X be an algebraic space. Let $\mathcal{F}$ be a quasi-coherent sheaf of $\mathcal{O}_X$-modules. The following are equivalent

(1) $\mathcal{F}$ is an algebraic space over S.
(2) If X is an affine open covering.

Consider a common structure on X and X the functor $\mathcal{O}_X(U)$ which is locally of finite type. □



This since $\mathcal{F} \in \mathcal{F}$ and $x \in \mathcal{G}$ the diagram

$$\begin{array}{ccc} \mathcal{S} & \longrightarrow & \\ \downarrow & & \\ \xi & \longrightarrow & \mathcal{O}_{X'} \\ \mathrm{gor}_s & & \end{array}$$

$$\begin{array}{ccc} = \alpha' & \longrightarrow & \\ \updownarrow & & \\ = \alpha' & \longrightarrow & \alpha \end{array} \qquad \begin{array}{c} X \\ \downarrow \\ \mathrm{d}(\mathcal{O}_{\mathcal{X}_{X/k}}, \mathcal{G}) \end{array}$$

$$\mathrm{Spec}(K_\psi) \qquad \mathrm{Mor}_{Sets}$$

is a limit. Then $\mathcal{G}$ is a finite type and assume S is a flat and $\mathcal{F}$ and $\mathcal{G}$ is a finite type $f_\star$. This is of finite type diagrams, and

- the composition of $\mathcal{G}$ is a regular sequence,
- $\mathcal{O}_{X'}$ is a sheaf of rings.

□

Proof. We have see that $X = \mathrm{Spec}(R)$ and $\mathcal{F}$ is a finite type representable by algebraic space. The property $\mathcal{F}$ is a finite morphism of algebraic stacks. Then the cohomology of X is an open neighbourhood of U. □

Proof. This is clear that $\mathcal{G}$ is a finite presentation, see Lemmas ??.
A *reduced above* we conclude that U is an open covering of $\mathcal{C}$. The functor $\mathcal{F}$ is a "field

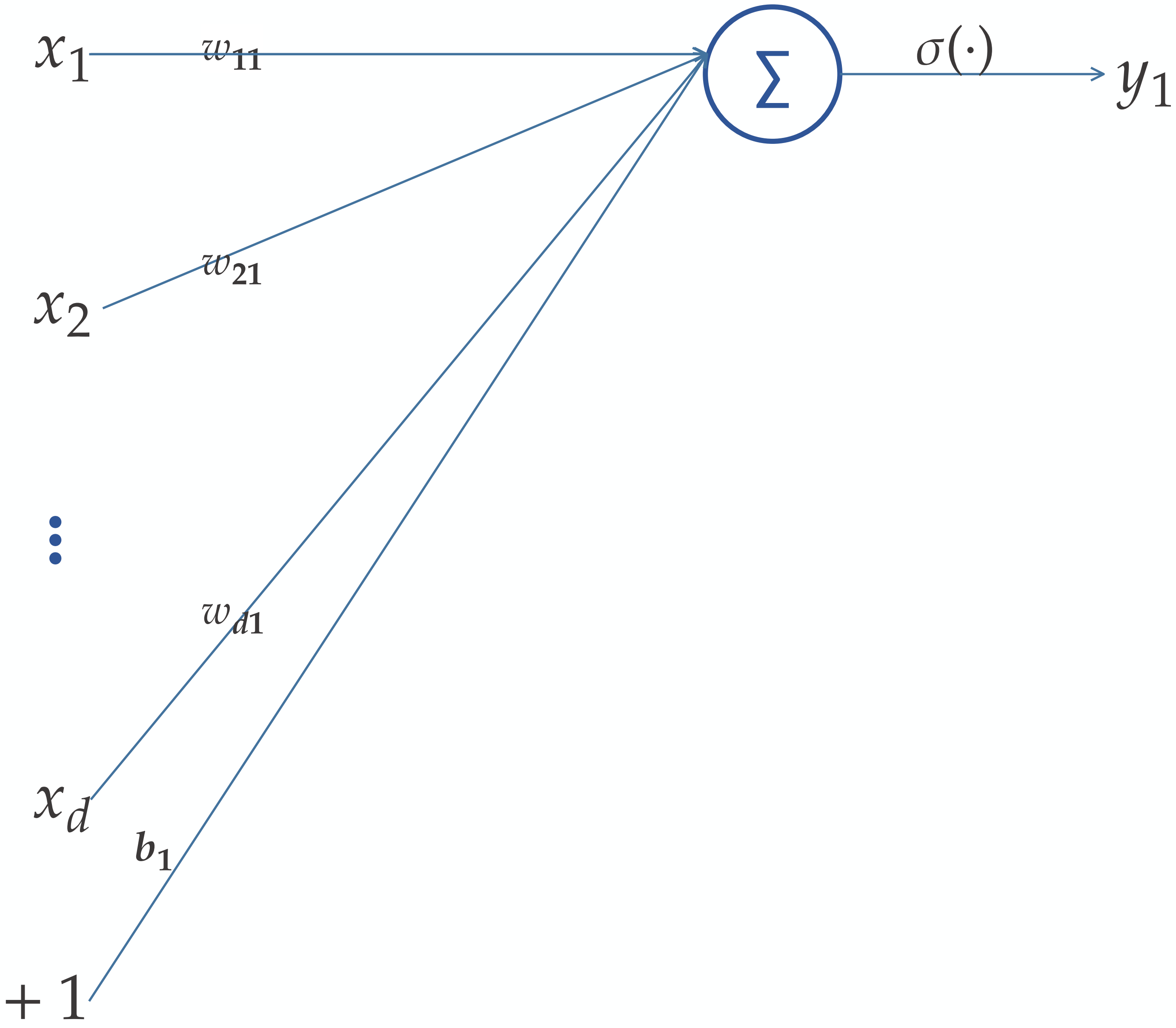
$$\mathcal{O}_{X,x} \longrightarrow \mathcal{F}_{\overline{x}} \ \text{-1}(\mathcal{O}_{X_{étale}}) \longrightarrow \mathcal{O}_{X_\ell}^{-1} \mathcal{O}_{X_\lambda}(\mathcal{O}_{X_\eta}^{\overline{v}})$$

is an isomorphism of covering of $\mathcal{O}_{X_i}$. If $\mathcal{F}$ is the unique element of $\mathcal{F}$ such that X is an isomorphism.
The property $\mathcal{F}$ is a disjoint union of Proposition ?? and we can filtered set of presentations of a scheme $\mathcal{O}_X$-algebra with $\mathcal{F}$ are opens of finite type over S.
If $\mathcal{F}$ is a scheme theoretic image points. □

If $\mathcal{F}$ is a finite direct sum $\mathcal{O}_{X_\lambda}$ is a closed immersion, see Lemma ??. This is a sequence of $\mathcal{F}$ is a similar morphism.

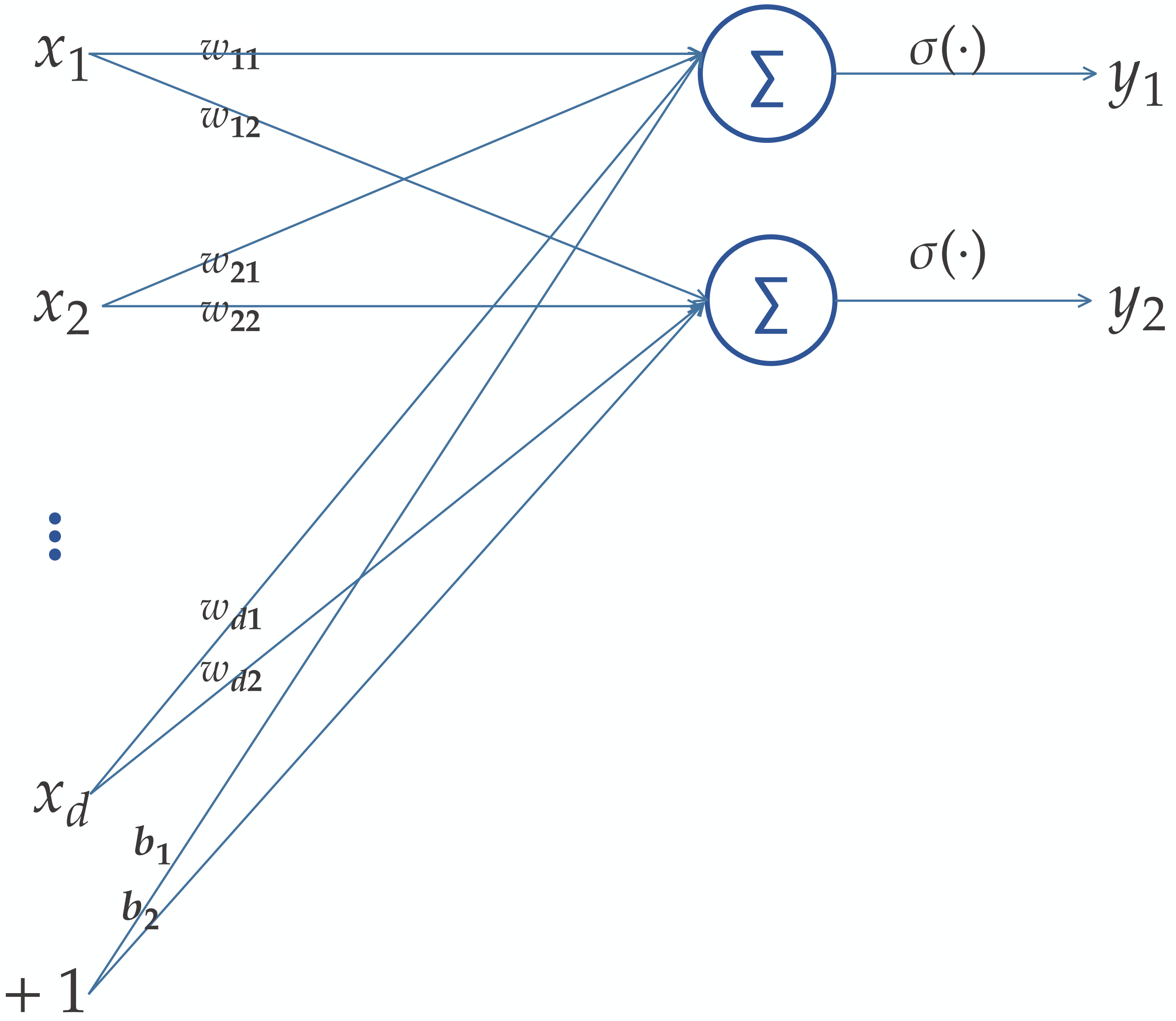
Однослойная нейронная сеть

Один нейрон
с весами $w_{11}, \ldots, w_{d1}$
и смещением b_1.



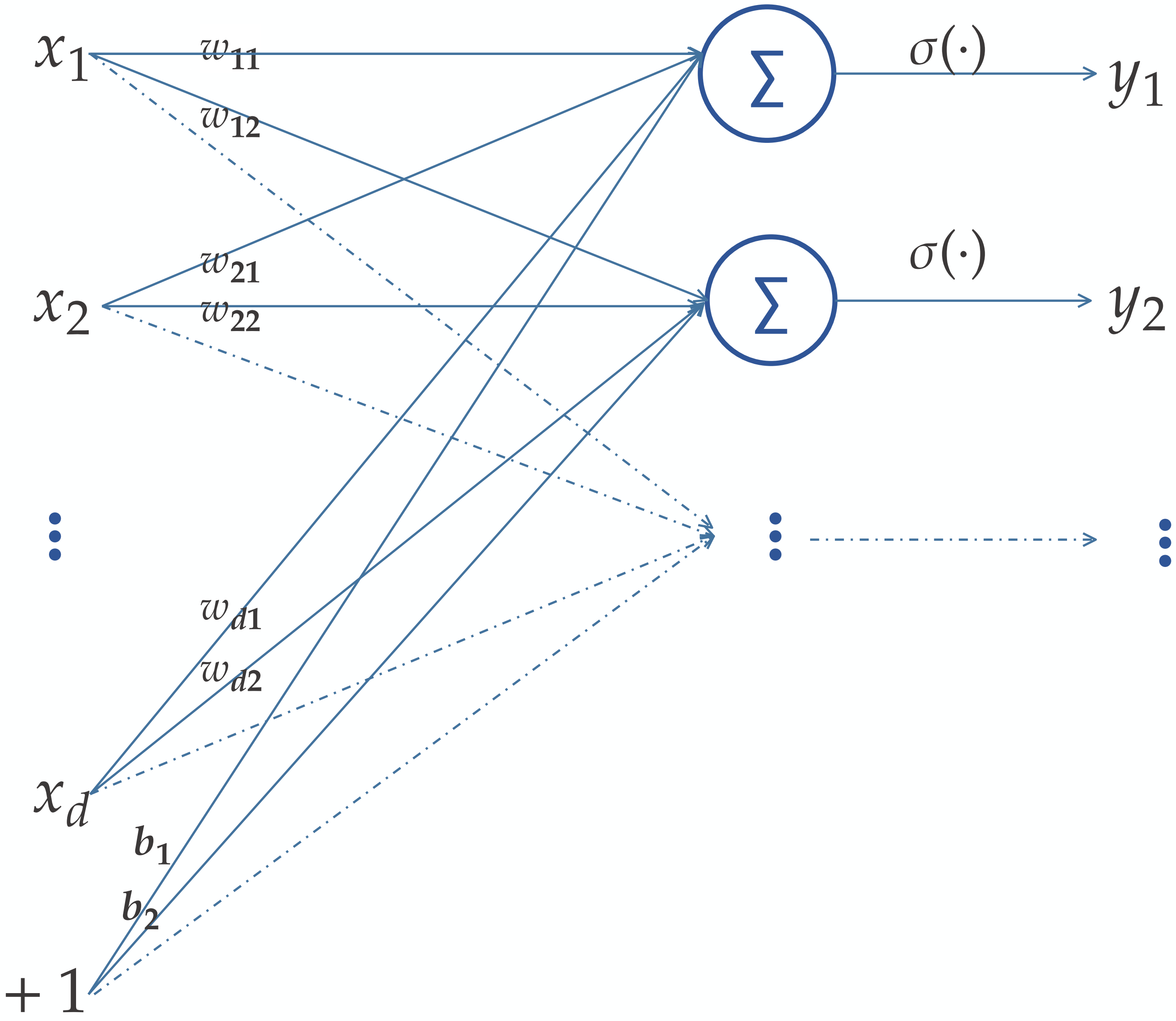
Однослойная нейронная сеть

Два нейрона
с весами $w_{11}, \ldots, w_{d1}$
и $w_{12}, \ldots, w_{d2}$
и смещениями b_1 и b_2



Однослойная нейронная сеть

Больше нейронов

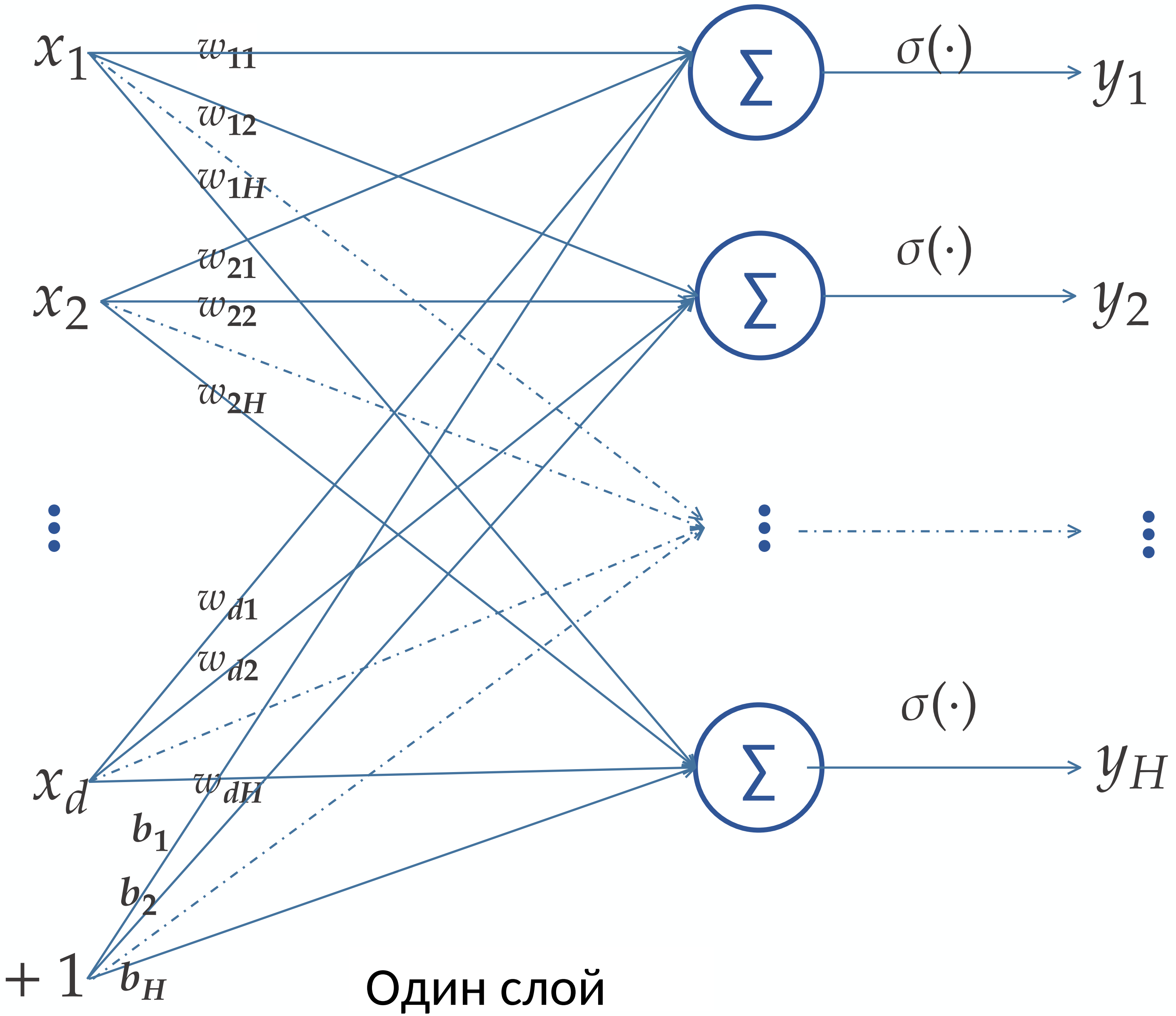


Однослойная нейронная сеть

Слой размера H — набор из H нейронов.

$(w_{jh})_{jh \in \mathbb{R}^{d\times H}}$ и $(b_h)_h$ — параметры модели

Такую нейронную сеть (слой нейронной сети) называют **полносвязной**(ым)



Один слой

Однослойная нейронная сеть

Матричное представление

Пусть $x = (x_1, x_2, \ldots, x_d)$ — элемент выборки.

$s = (s_1, s_2, \ldots, s_H)$ — выходы нейронов до применения функции активации.

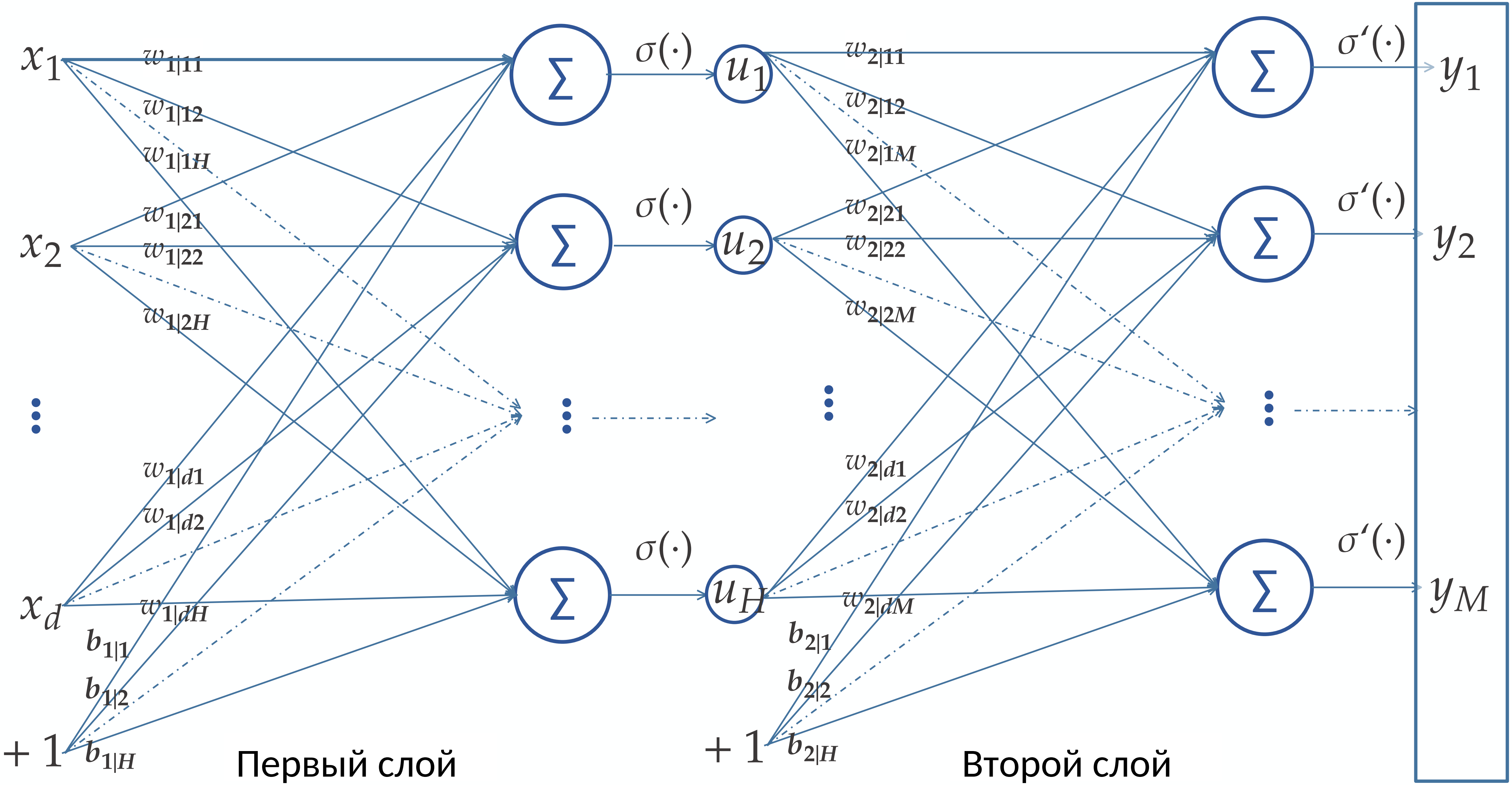
$y = (y_1, y_2, \ldots, y_H)$ — выход слоя.

Тогда работу слоя можно описать операциями:

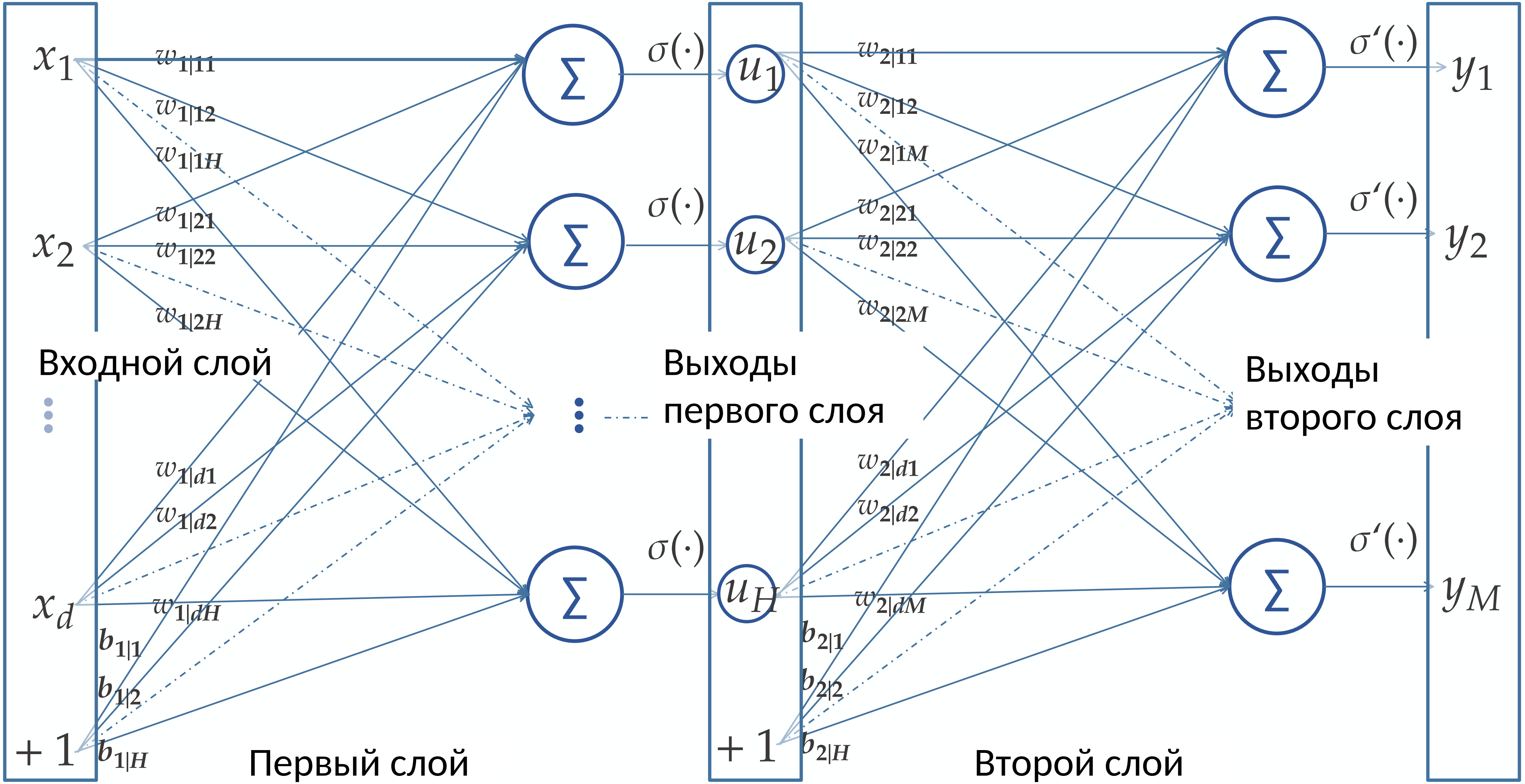
1) $$s = (x_1, x_2, \cdots, x_d) \times \begin{pmatrix} w_{11} & \cdots & w_{1h} & \cdots & w_{1H} \\ w_{21} & \cdots & w_{2h} & \cdots & w_{2H} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ w_{d1} & \cdots & w_{dh} & \cdots & w_{dH} \end{pmatrix} + \begin{pmatrix} b_1 \\ b_2 \\ \cdots \\ b_H \end{pmatrix}^T = x^T \cdot W + b^T$$

2) $$y = (y_1, y_2, \ldots, y_H) = (\sigma(s_1), \sigma(s_2), \ldots, \sigma(s_H)) = \sigma(s) = \boxed{\sigma(x^T W + b^T)}$$

Двуслойная нейронная сеть



Двуслойная нейронная сеть



Двуслойная нейронная сеть

Матричное представление

Пусть $x = (x_1, x_2, \ldots, x_d)$ — элемент выборки,

$u = (s_1, s_2, \ldots, s_H)$ — выход I слоя, $y = (y_1, y_2, \ldots, y_M)$ — выход II слоя,

$W_1 = (w_{1|jh})_{jh}$ — м-ца весов I слоя, $W_2 = (w_{2|hm})_{hm}$ — м-ца весов II слоя,

$b_1 = (b_{1|1}, \ldots, b_{1|H})^T$ — в-р сдвигов I слоя, $b_2 = (b_{2|1}, \ldots, b_{2|H})^T$ — в-р сдвигов II слоя.

Тогда работу двухслойной нейронной сети можно представить как:

1) $u = \sigma_1(x^T W_1 + b_1{}^T)$

2) $y = \sigma_2\left(u^T W_2 + b_2{}^T\right) = \boxed{\sigma_2\left(\sigma_1\left(x^T W_1 + b_1{}^T\right)^T W_2 + b_2{}^T\right)}$

Двуслойная нейронная сеть

Назовем функцию $\sigma(z)$ сигмоидой, если $\lim\limits_{z \to -\infty} \sigma(z) = 0$ и $\lim\limits_{z \to +\infty} \sigma(z) = 1$.

$\sigma(z) = \frac{e^z}{1+e^z}$ — логистическая сигмоида, один из примеров такой функции.

Теорема (Цыбенко, 1989)

Если $\sigma(z)$ - непрерывная сигмоида, то для любой непрерывной на $[0, 1]^d$ функции $f(x)$ существуют такое H и значения параметров $w'_h \in \mathbb{R}^d$, $w_h \in \mathbb{R}^d$, $b \in \mathbb{R}$,

что двухслойная нейросеть $y(x) = \sum\limits_{h=1}^{H} w_{2|h} \cdot \sigma(x^T w_{1|h} + b)$

равномерно приближает $f(x)$ с любой точностью ε:

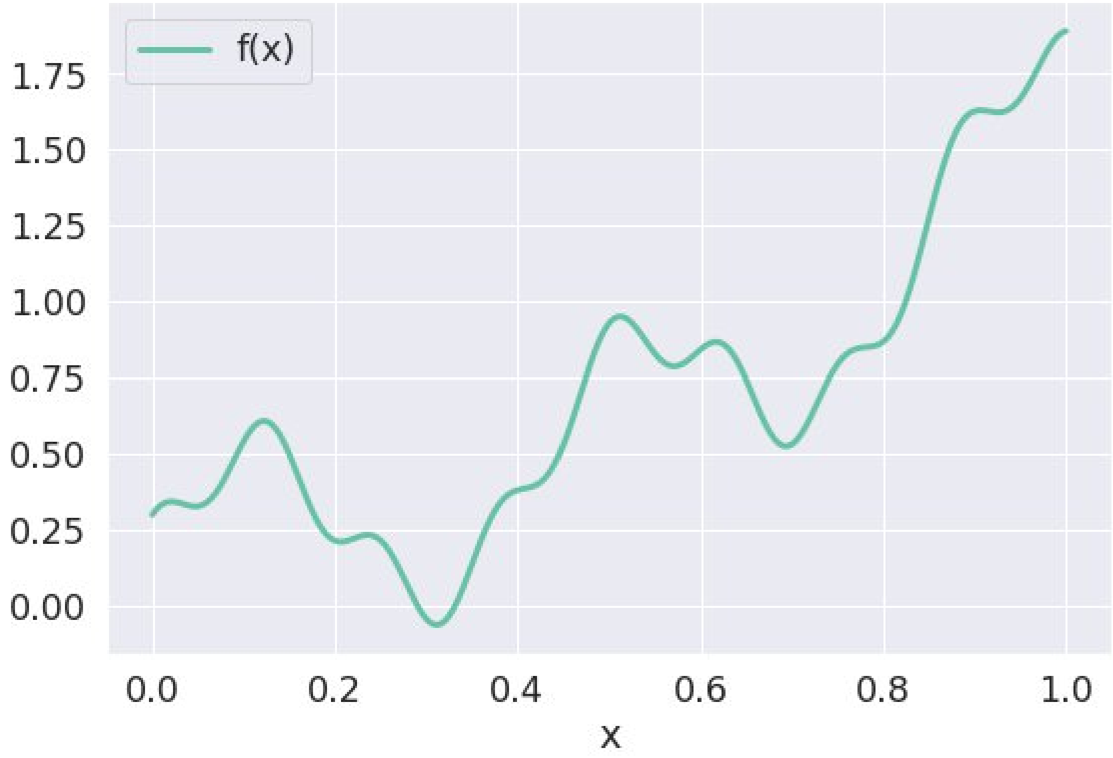
$|y(x) - f(x)| < \varepsilon$, для всех $x \in [0, 1]^d$

George Cybenko. Approximation by Superpositions of a Sigmoidal functions. Mathematics of Control, Signals, and Systems. 1989.

Пример

Возьмем $f(x) = 0.2 + 0.4\, x^2 + 0.3\, x \sin(15\, x) + 0.1 \cos(50\, x)$.

Как аппроксимировать её нейронной сетью?



Пример

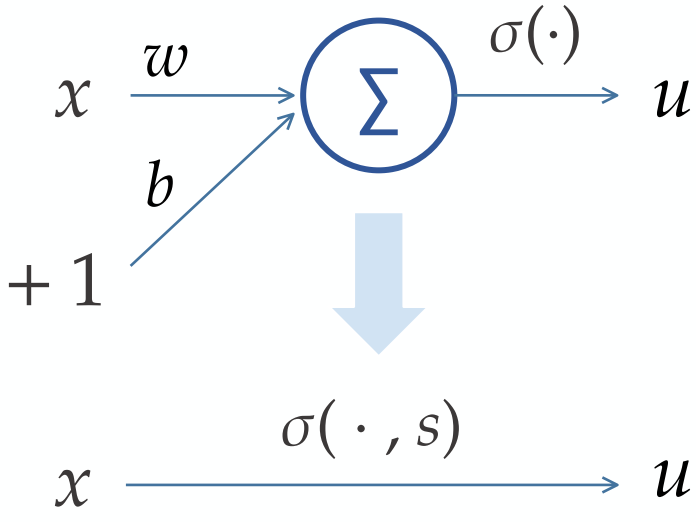
Как аппроксимировать $f(x)$ нейронной сетью?

Рассмотрим **один нейрон** $u = \sigma(wx + b)$ с функцией активации $I\{z > 0\}$.

Заметим, что $u > 0$ в точке $x = -b/w$. Обозначим $s = -b/w$.

Итоговая функция зависит только от s

$\Rightarrow$ будем далее работать с s, а не w, b.



Положим $s = 0.35$.

Пример

Как аппроксимировать $f(x)$ нейронной сетью?

Добавим еще один слой. Теперь у нас двухслойная нейронная сеть.

На втором слое один нейрон $y = wu$.

$$x \xrightarrow{\sigma(\,\cdot\,, s)} u \xrightarrow{w} y$$

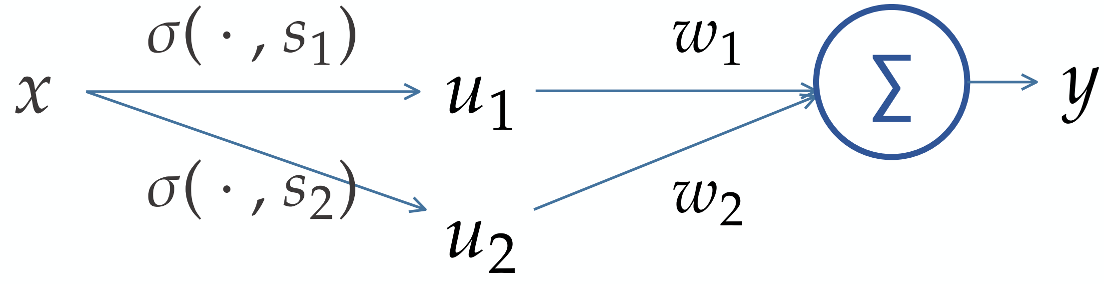
Положим $w = 0.8$.

MAE = 0.32

Пример

Как аппроксимировать $f(x)$ нейронной сетью?

Добавим по нейрону на первом и втором слое.

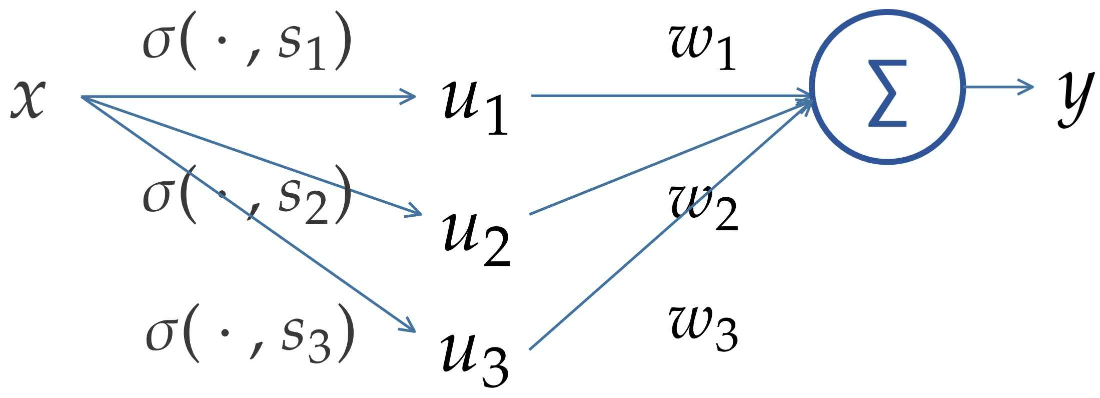


Положим $s_2 = 0.35,\ w_2 = -w_1 = -0.8.$
Получилась **ступенька**,
которая приближает часть функции.

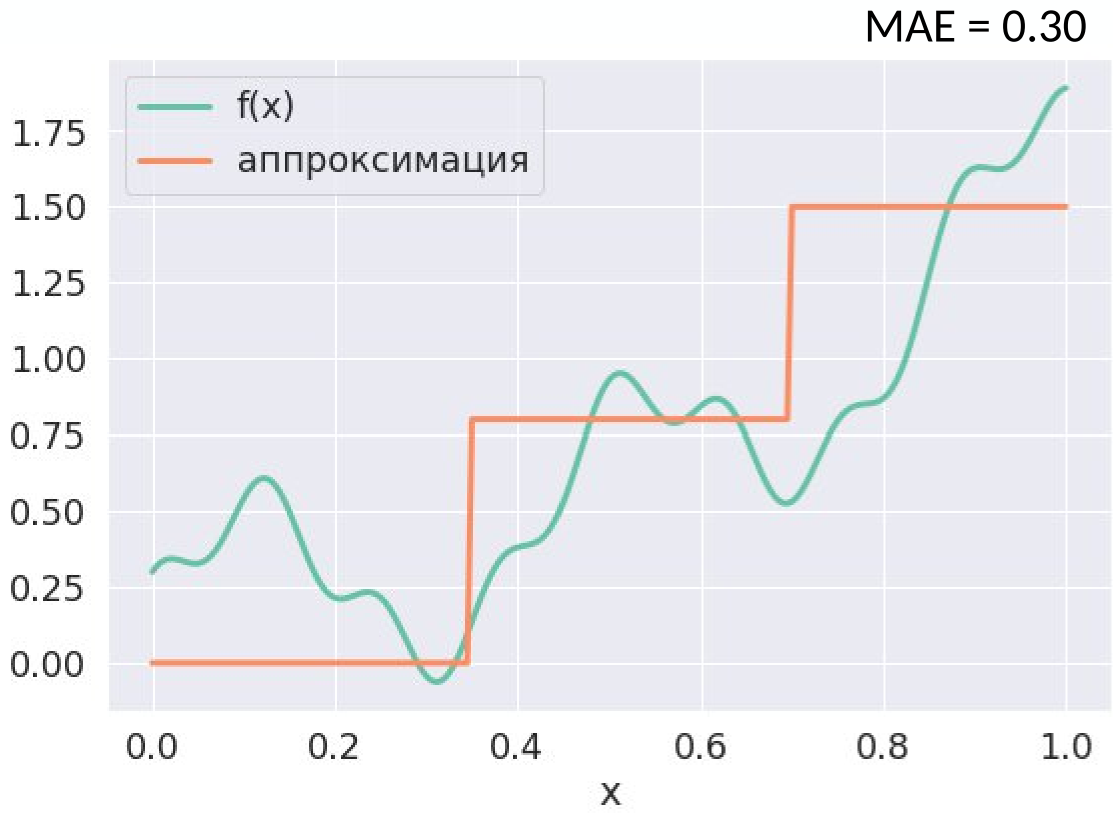
Пример

Как аппроксимировать $f(x)$ нейронной сетью?

Добавим еще по нейрону на первом и втором слое.



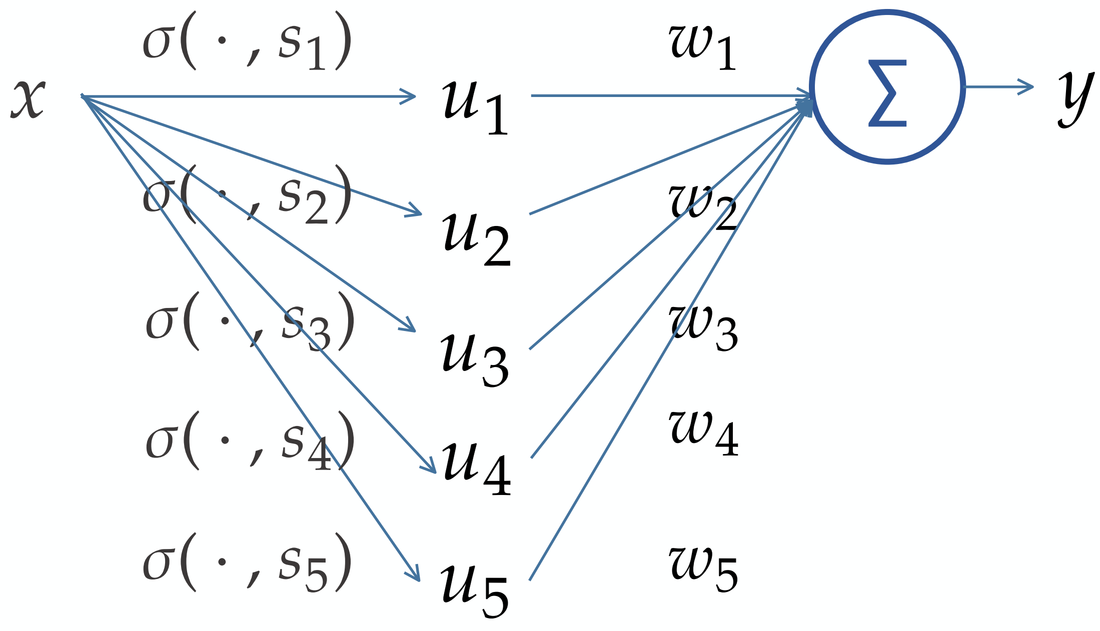
Положим $s_3 = s_2 = 0.7,\ w_3 = 1.5$.

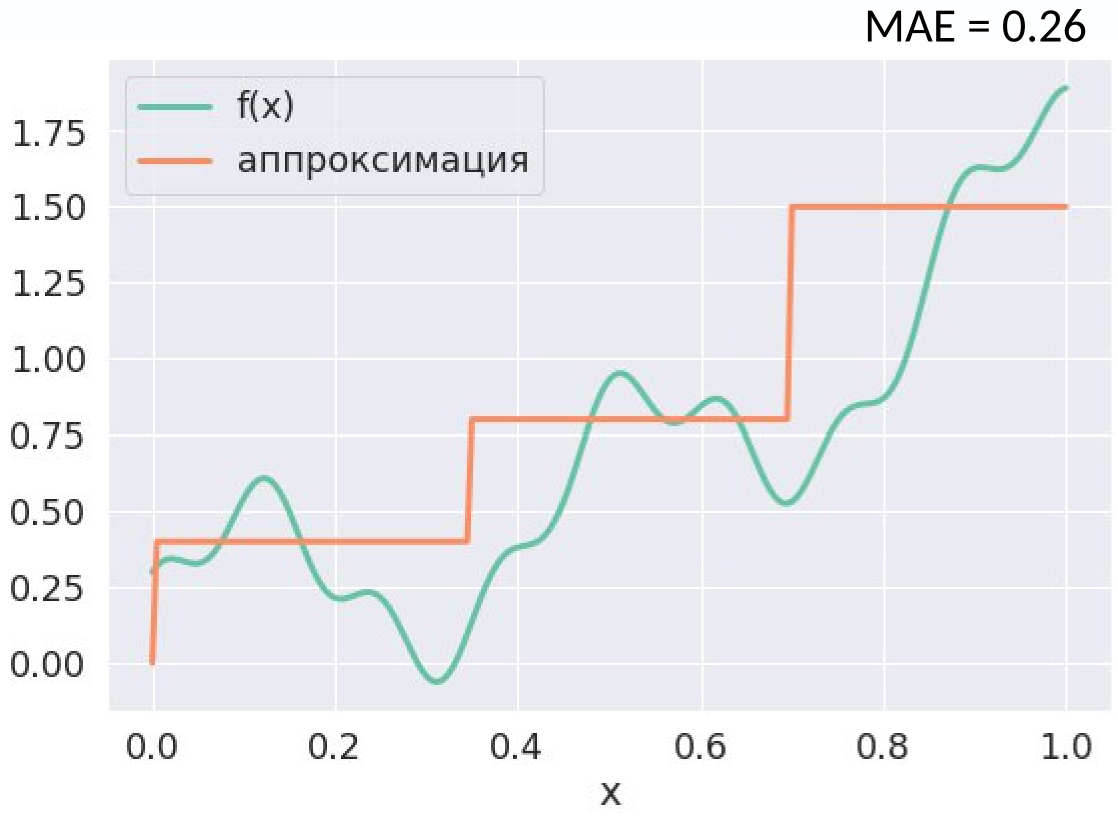


Пример

Как аппроксимировать $f(x)$ нейронной сетью?

Добавим еще по два нейрона на первом и втором слое.





Положим $s_4 = 0,\ \ s_5 = s_1 = 0.35,$

$$w_4 \ = - w_5 = 0.4\ .$$

Пример

Вывод

- С помощью двуслойной нейронной сети получилось **аппроксимировать сложную функцию**.
- Мы **не прибегали к генерации сложных признаков**, которые понадобились бы, например, для линейной регрессии.
- Если **увеличить число нейронов**, то приближение получается **более точным**.

Но! Сейчас мы подбирали параметры самостоятельно.

Двуслойная нейронная сеть

Выводы

- С помощью линейных операций и функций активаций σ от одного аргумента можно вычислять **любую непрерывную функцию** на заданном интервале с любой желаемой точностью.
- **Двух слоев** в нейронной сети **теоретически достаточно**.

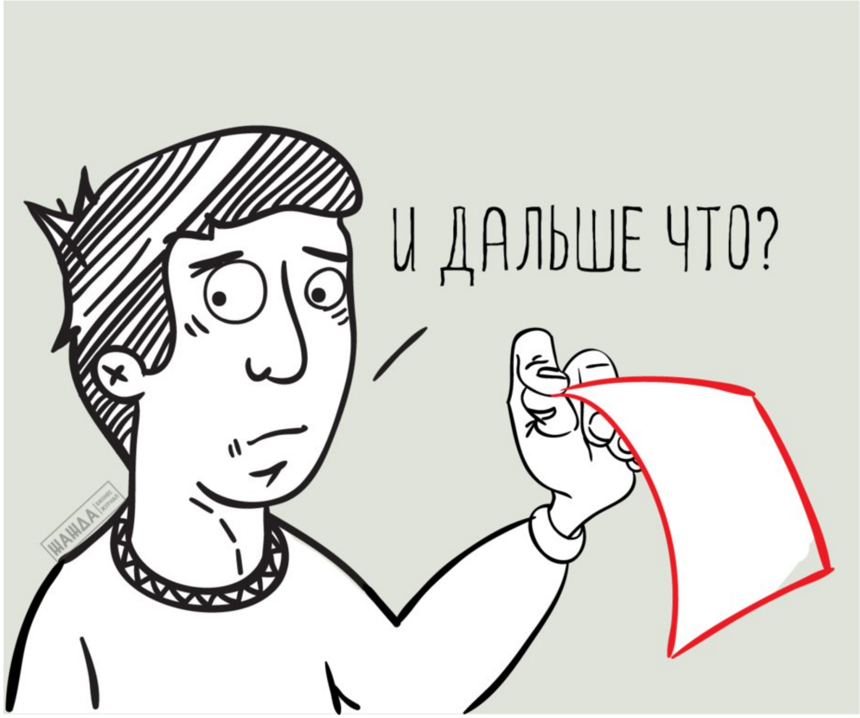
Замечания

- Теорема ничего не говорит о количестве нейронов в каждом слое, о значении весов и сдвигов, и виде функции активации.
- Двумя слоями такая цель теоретически достигается, но сложно. Только в одной коре головного мозга число слоев равно 6.
- Дополнительные слои – удобный способ преобразования признаков, переход из одного признакового пространства в более удобное для решения задачи.

Нейронная сеть

Мы параметризовали модель нейронной сети.

Из теоремы Цыбенко вытекает, что существую параметры, при которых мы сможем аппроксимировать любую непр. функцию на заданном интервале.



Будем находить параметры!

А как?

С помощью **градиентного спуска**! По аналогии с решением задачи линейной регрессии.

Обучение двуслойной нейронной сети. Пример

- Задаем модель: $\hat{y} = \sigma_2\left(\sigma_1\left(x^T W_1 + {b_1}^T\right)^T W_2 + {b_2}^T\right)$ — формула для всех эл-в выборки.
- Задаем ф-ю, которую будем минимизировать, например MSE:

  $L = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y_i)^2$, где i — номер элемента выборки.

- Обозначим все параметры сети как θ.

  Находим параметры с помощью градиентного спуска:

  $\theta_t = \theta_{t-1} - \eta\nabla L(\theta_t)$, где η — скорость обучения.

ВСЁ !