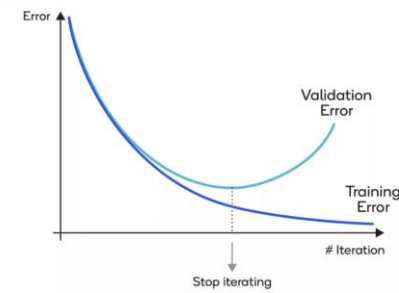
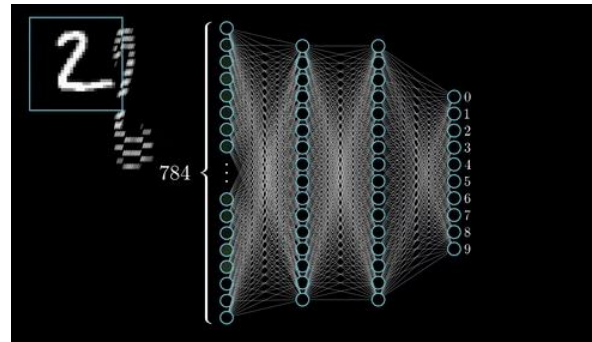
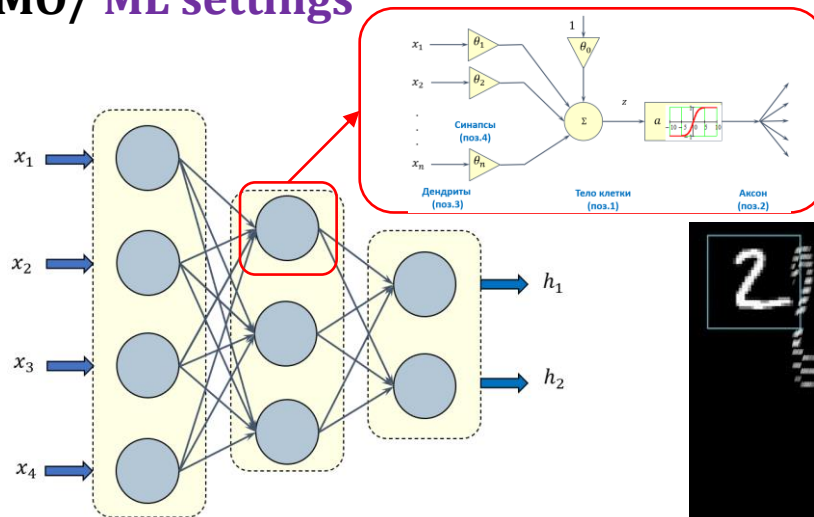


Лекция 2. Искусственные нейронные сети (ИНС) и обучение с учителем / Lecture 2. Artificial Neural Networks (ANNs) and Supervised Learning

1. Напоминание / Contents of the previous lecture
2. Естественные и искусственные нейронные сети (ИНС)/ Natural and artificial neural networks (ANNs)
3. Прямые вычисления в ИНС / Forward propagation
4. Обратные вычисления в ИНС/ Back propagation
5. Настройка МО/ ML settings



Напоминание / Contents of the previous lecture



Машинное обучение / Machine Learning (ML):
задача «З», в ходе решения которой программа обучается из опыта «О» и повышает меру качества «К»

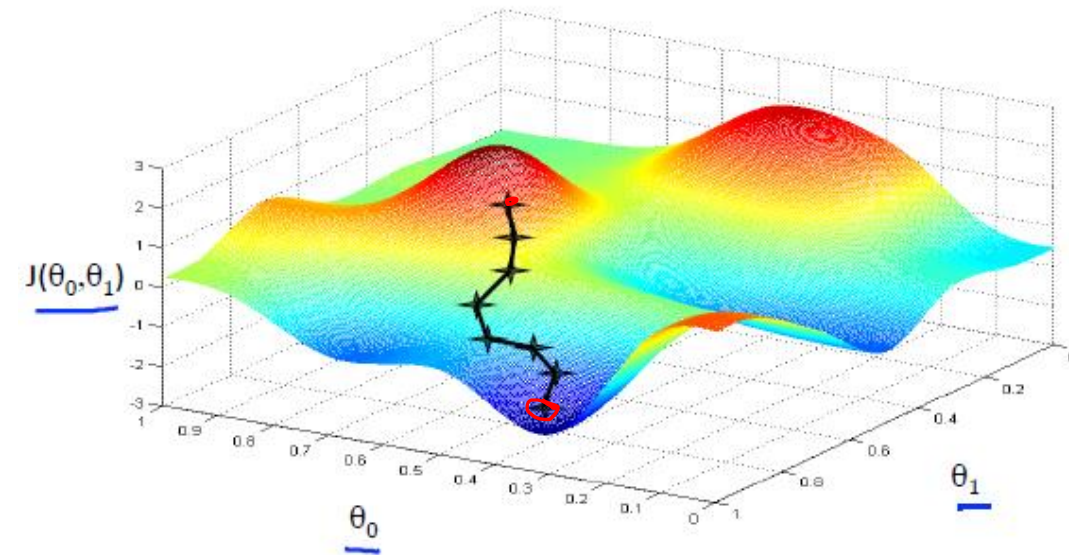
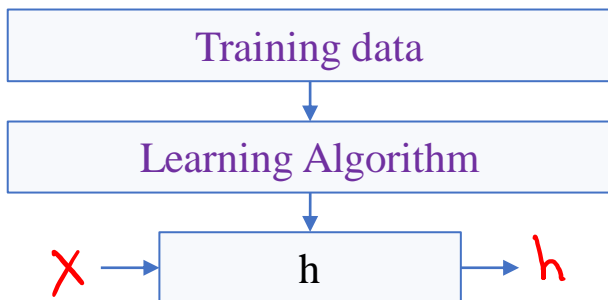
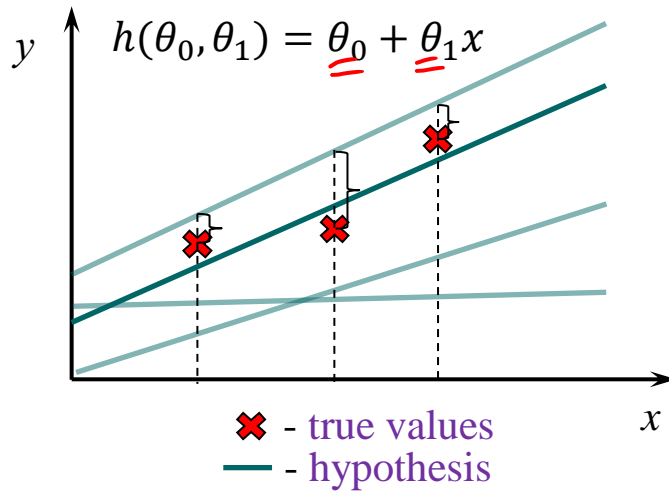
The main idea is to modeling the relationship between two sets: a scalar response (y) and independent variable (x) by minimizing the cost function:

$$X = \begin{pmatrix} x^{(1)} \\ \vdots \\ x^{(m)} \end{pmatrix}; \bar{Y} = \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{pmatrix}; \tilde{X} = \begin{pmatrix} 1 & x^{(1)} \\ \vdots & \vdots \\ 1 & x^{(m)} \end{pmatrix}; \Theta = \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$$
$$H = \tilde{X} \Theta$$

using gradient descent method, for instance.

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h^{(i)} - y^{(i)})^2 \Rightarrow \min.$$

$-\alpha \nabla J$



Gradient descent intuition

Напоминание / Contents of the previous lecture

Логистическая р. / Logistic r.

1 перем. x / 1 variable x

Мн.перем. x_k / Multiple var. x_k
($k = 1 \dots n$)

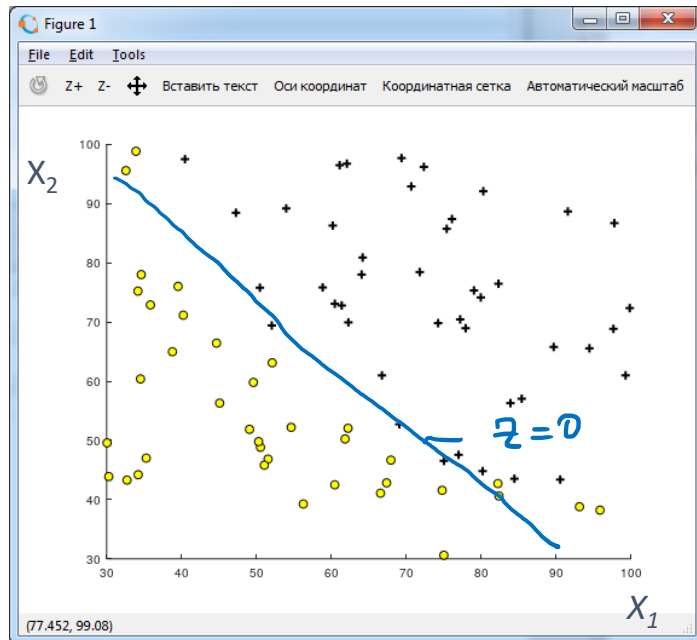
Полином. / Polynomial

$k = 2$

$$X = \begin{pmatrix} \begin{pmatrix} x_1^{(1)} & x_2^{(1)} \\ \vdots & \vdots \\ x_1^{(m)} & x_2^{(m)} \end{pmatrix} \end{pmatrix}; Y = \begin{pmatrix} \begin{pmatrix} y^{(1)} \\ \vdots \\ y^{(m)} \end{pmatrix} \end{pmatrix}; \rightarrow X = \begin{pmatrix} \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} \\ \vdots & \vdots & \vdots \\ 1 & x_1^{(m)} & x_2^{(m)} \end{pmatrix} \end{pmatrix}; \Theta = \begin{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix} \end{pmatrix}; \underline{Z = X\Theta}; \rightarrow h(z) = \frac{1}{1+e^{-z}}.$$

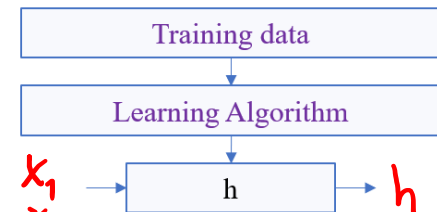
$$J(\Theta) = -\frac{1}{m} \sum_{i=1}^m (y^{(i)} \ln(h^{(i)}) + (1 - y^{(i)}) \ln(1 - h^{(i)})) \Rightarrow \min.$$

$$z = \theta_0 + \theta_1 x_1 + \theta_2 x_2 = 0 \quad x_2 = -\frac{\theta_0}{\theta_2} - \frac{\theta_1}{\theta_2} x_1$$

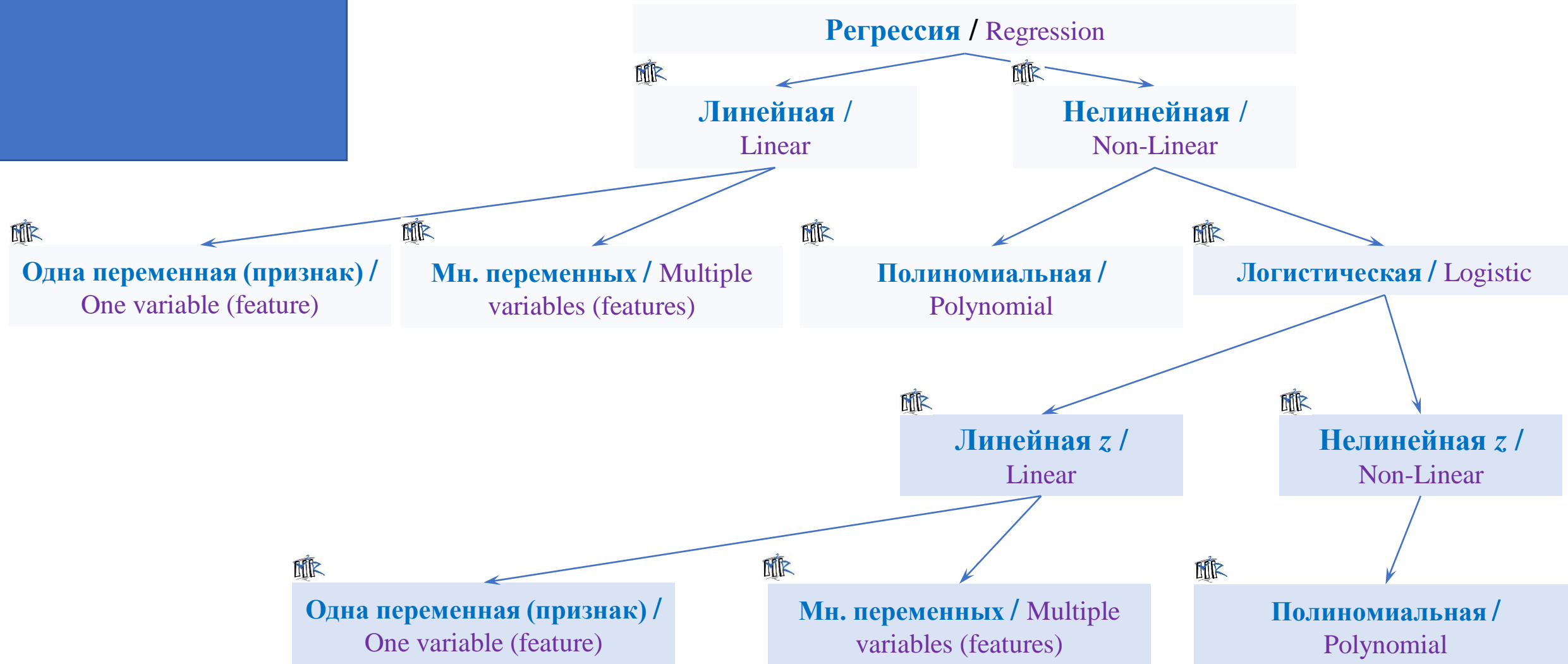


Алгоритм поиска минимума функции качества.

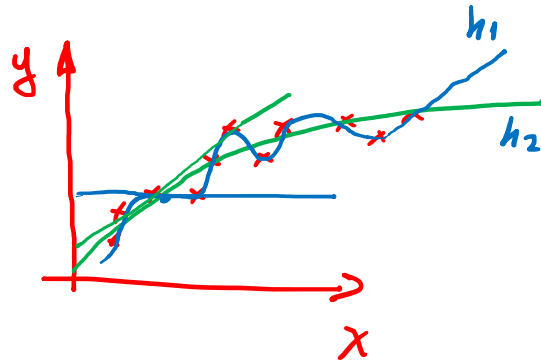
1. Задать начальные значения компонент матрицы Θ случайным образом
2. Рассчитать $H = X\Theta$ и $\nabla J = \frac{1}{m} X^T (H - Y)$.
3. Найти новые значения компонент матриц Θ^H : $\underline{\Theta^H = \Theta^C - \alpha \nabla J}$.
4. Повторять пункты 2-3 до выполнения одного из условий: $J^H - J^C < \delta$, #итер. $> N_{max}$.
5. Вывод результатов: Θ .



Напоминание / Contents of the previous lecture



Напоминание / Contents of the previous lecture



$$\nabla J = \frac{1}{m} X^T (H - Y); \quad \nabla J = 0$$

$$X^T H - X^T Y = 0; \quad H = X \Theta;$$

$$X^T X \Theta = -X^T Y.$$

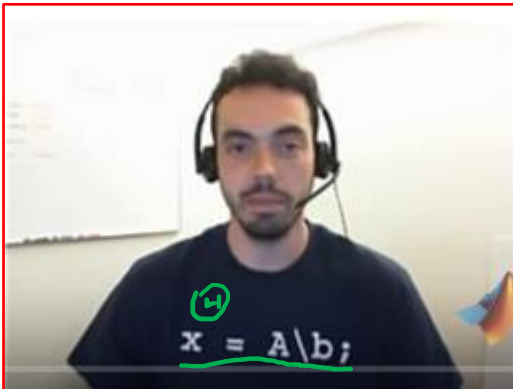
$$(X^T X)^{-1} X^T X \Theta = -(X^T X)^{-1} X^T Y$$

$$\Theta = -(X^T X)^{-1} X^T Y$$

①

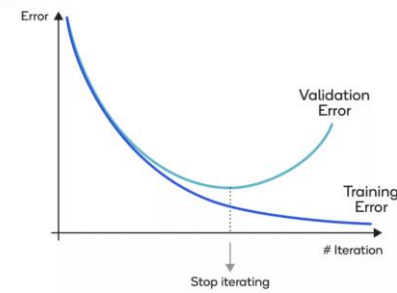
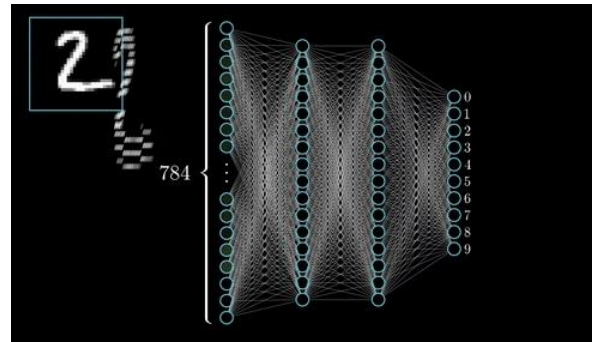
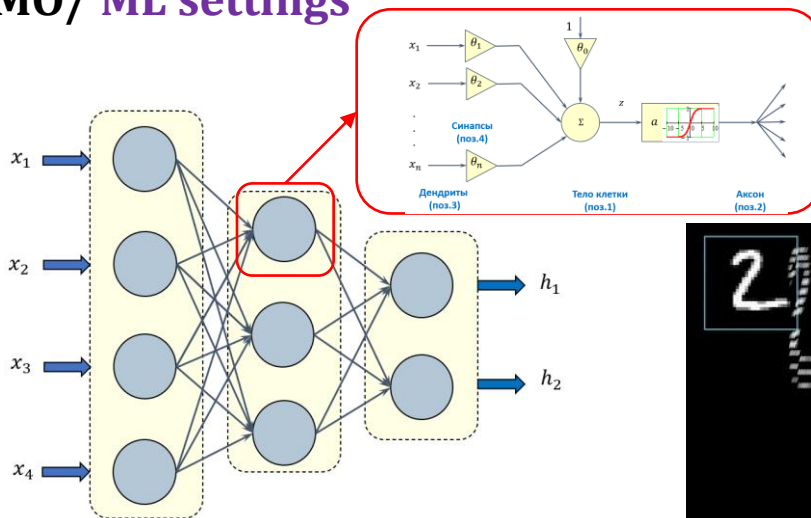
Вопросы:

1. Почему не рекомендуется дифференцировать функции, полученные в результате регрессионного анализа данных?
2. Почему нельзя сокращать «d» в выражении «dy/dx»?
3. В каких случаях математическое ожидание не совпадает со средним арифметическим значением?
4. Каким образом можно улучшить метод градиентного спуска, чтобы находить с его помощью глобальные минимумы, вместо локальных?
5. Можно ли рассмотренные задачи линейной регрессии решить аналитически, без применения метода градиентного спуска?
6. Почему при построении регрессионных моделей обычно не рекомендуется применение полиномов высоких степеней?



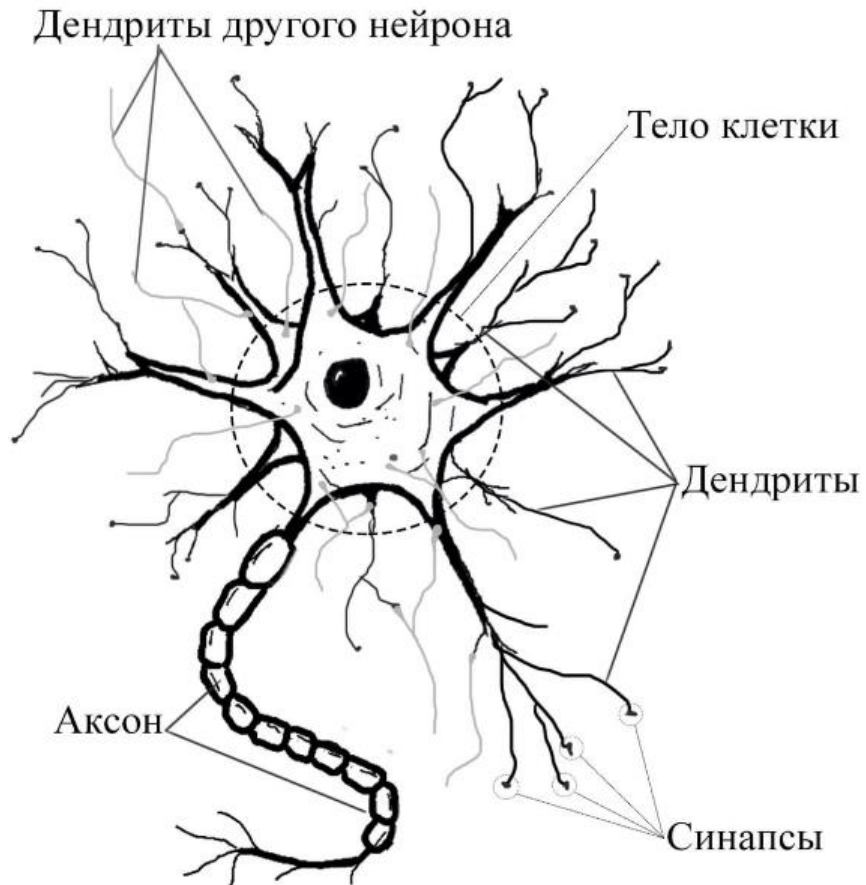
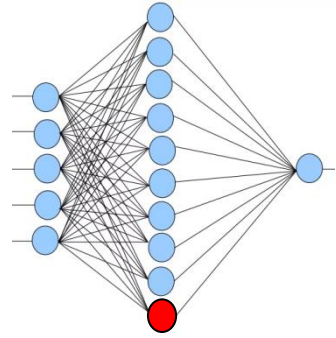
Лекция 2. Искусственные нейронные сети (ИНС) и обучение с учителем / Lecture 2. Artificial Neural Networks (ANNs) and Supervised Learning

1. Напоминание / Contents of the previous lecture
2. Естественные и искусственные нейронные сети (ИНС)/ Natural and artificial neural networks (ANNs)
3. Прямые вычисления в ИНС / Forward propagation
4. Обратные вычисления в ИНС/ Back propagation
5. Настройка МО/ ML settings

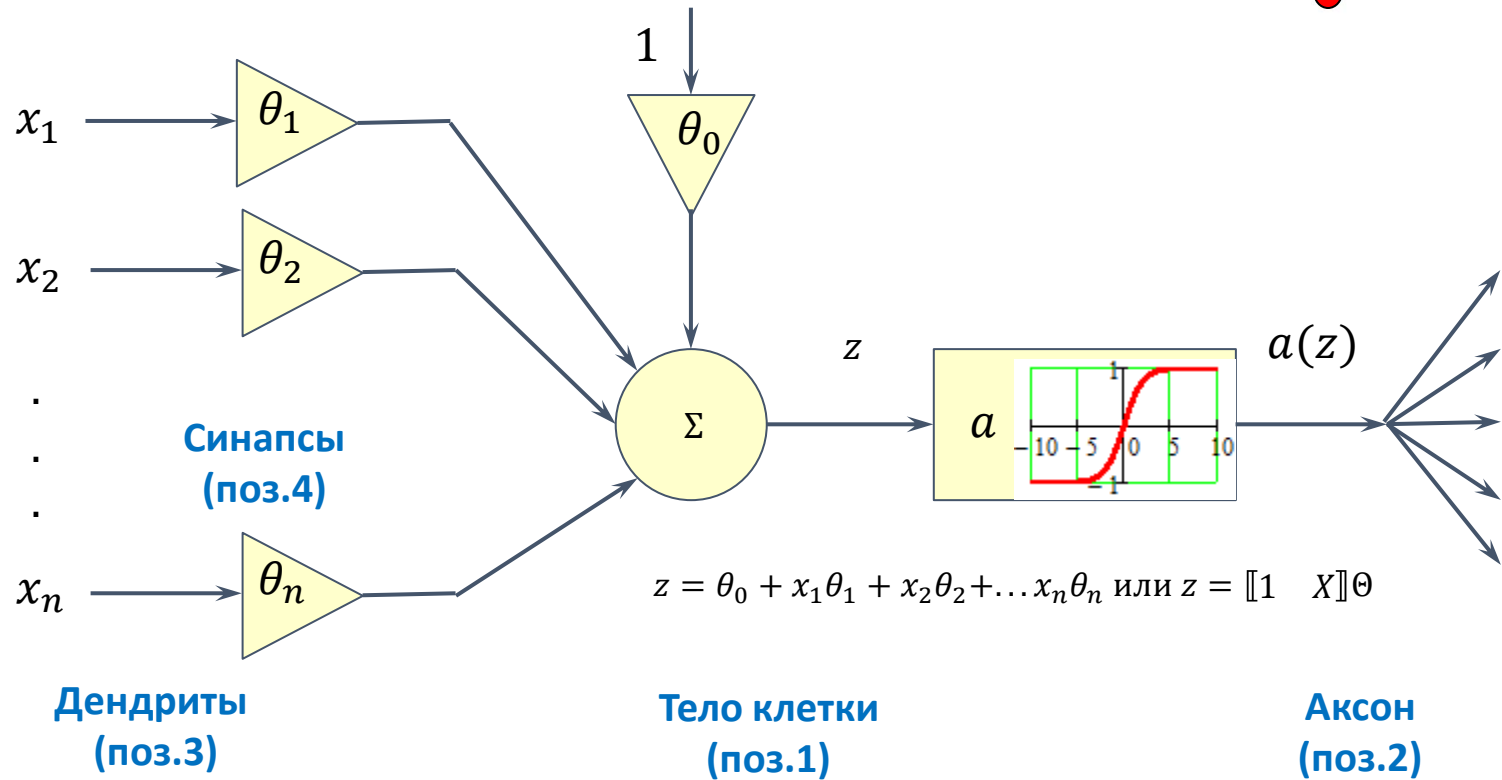


Естественные и искусственные нейронные сети / Natural and artificial neural networks

- 1) **Тело клетки** – обрабатывает информацию;
- 2) **аксон** – передает обработанную информацию другим нейронам;
- 3) **дендриты** – получают информацию от других нейронов;
- 4) **синапсы** – соединяют аксон и дендриты других нейронов.



Естественный нейрон

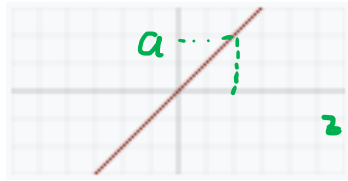


Искусственный нейрон

Естественные и искусственные нейронные сети / Natural and artificial neural networks

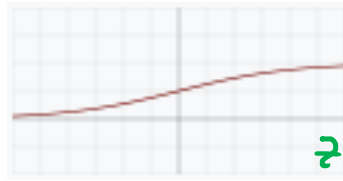
Функции активации / Activation functions

$$a = z$$



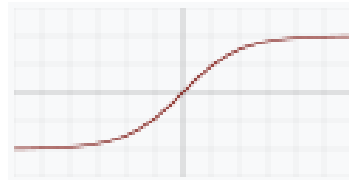
Identity

$$a = \frac{1}{1 + e^{-z}}$$



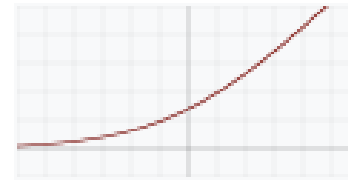
Logistic

$$a = \frac{e^z - e^{-z}}{e^z + e^{-z}}$$



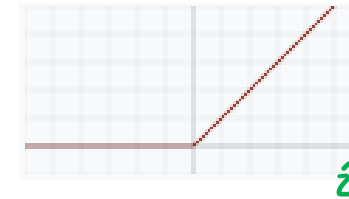
Hyperbolic tangent

$$a = \ln(1 + e^z)$$



Softplus

$$a = \max(0, z)$$



Rectified linear
Unit (ReLU)

$$a_j = \frac{e^{z_j}}{\sum_i e^{z_i}}$$

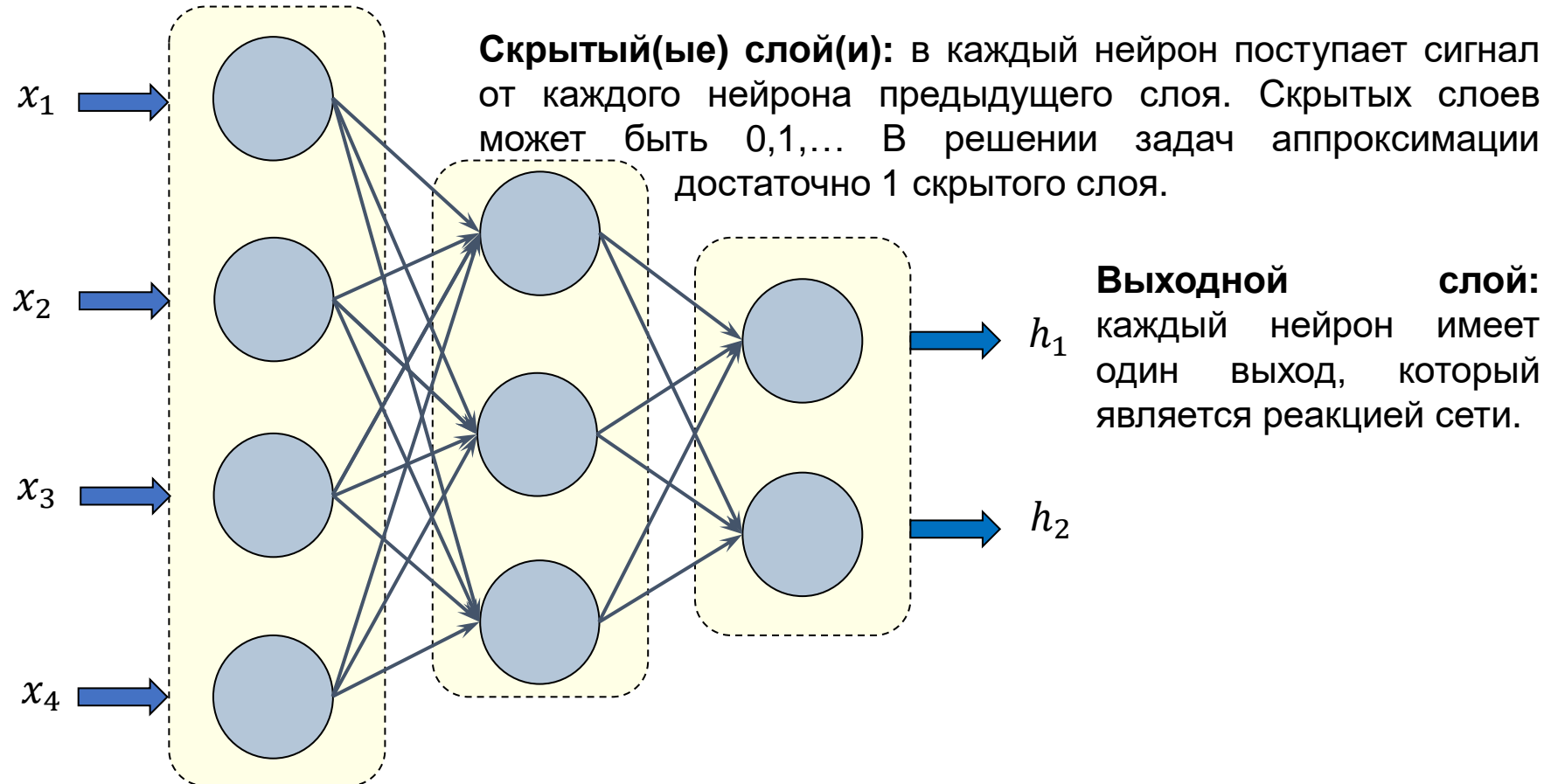


Softmax

Естественные и искусственные нейронные сети / Natural and artificial neural networks

Архитектура сетей прямого распространения / **Feed forward neural network**

Входной слой: каждый нейрон имеет ровно один вход от внешней среды.



Естественные и искусственные нейронные сети / Natural and artificial neural networks

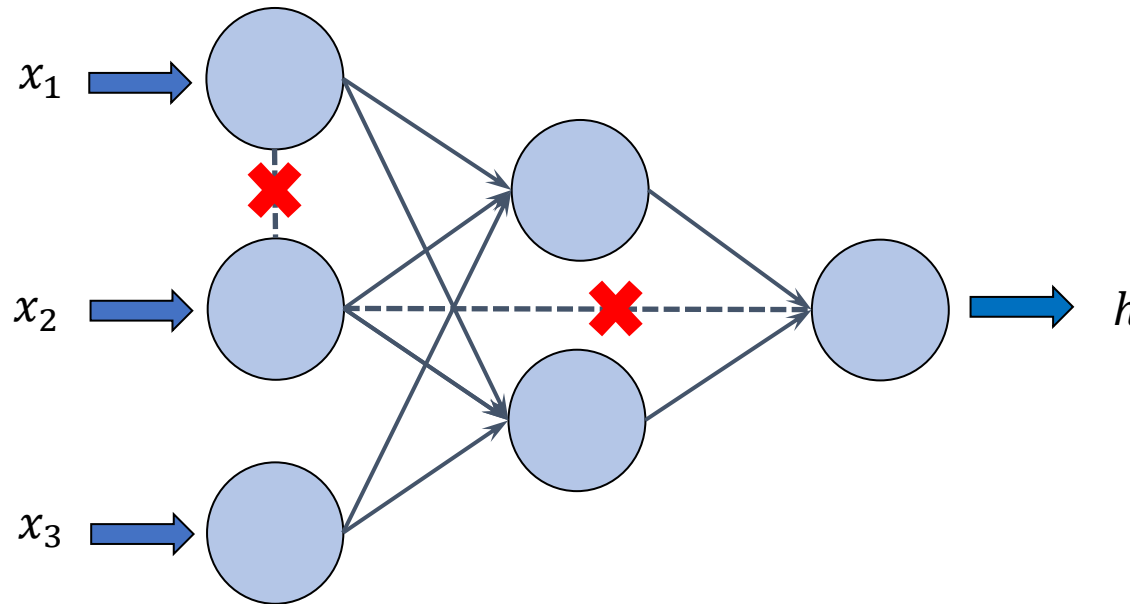
Архитектура сетей прямого распространения / Feed forward neural network

Нейроны слоя не связаны.

Нейроны передают информацию только нейронам следующего слоя.

Перепрыгивание через слои запрещено.

Настройка сети



Входы: x_i
Выход: h

Задаваемые и не варьируемые
параметры нейронной сети:

- кол-во входных нейронов;
- кол-во скрытых слоев;
- кол-во нейронов в скрытых слоях;
- количество выходных нейронов;
- активац. ф-я нейронов;
- ...

Направление потока информации

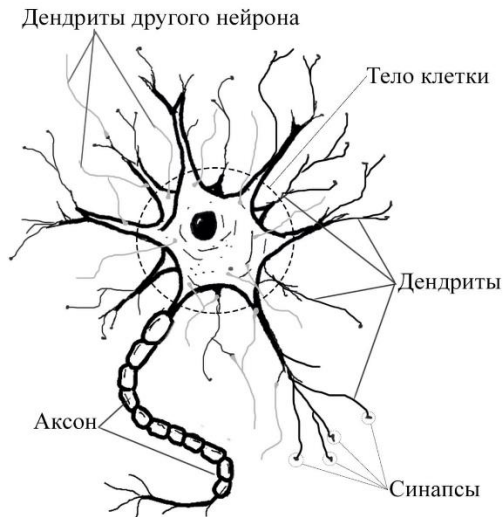
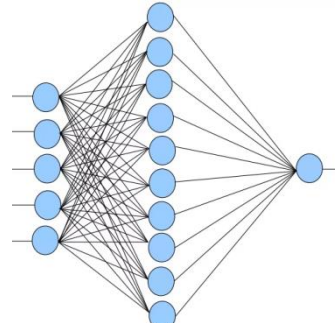


Варьируемые параметры нейронной сети:

- веса каждого соединения.

Естественные и искусственные нейронные сети / Natural and artificial neural networks

- 1) **Тело клетки** – обрабатывает информацию;
- 2) **аксон** – передает обработанную информацию другим нейронам;
- 3) **дендриты** – получают информацию от других нейронов;
- 4) **синапсы** – соединяют аксон и дендриты других нейронов.



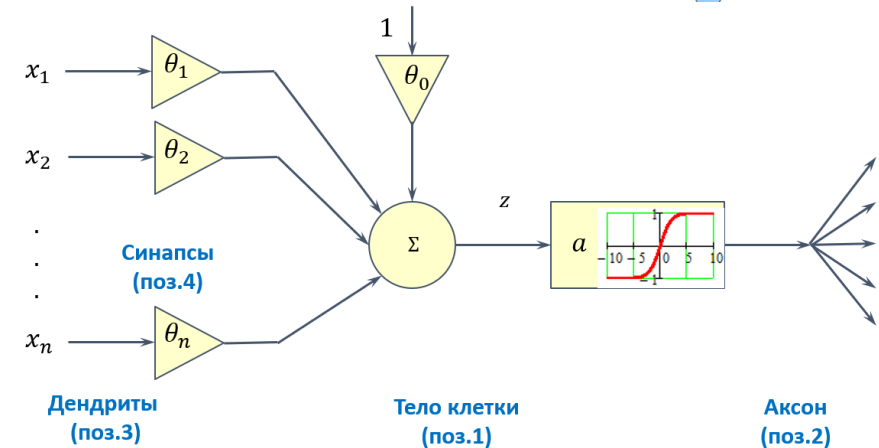
VS

Длительность 1 операции (с):

10^{-3}

10^{-9}

Архитектура нейронной сети:



В коре головного мозга порядка 10^9 нейронов и 10^{12} синаптических связей. Каждый нейрон связан с 10^4 соседних нейронов.

В ИНС порядка 10^2 нейронов и 10^4 синаптических связей. Каждый нейрон связан с 10^2 соседних нейронов.

Энергозатраты на выполнение 1 операции в секунду:

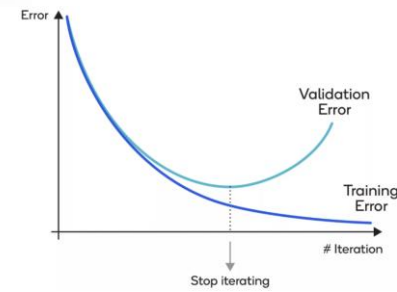
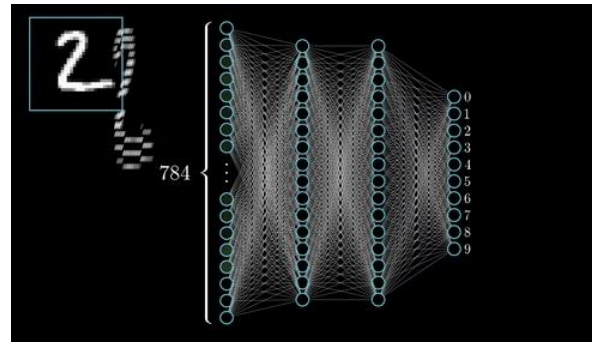
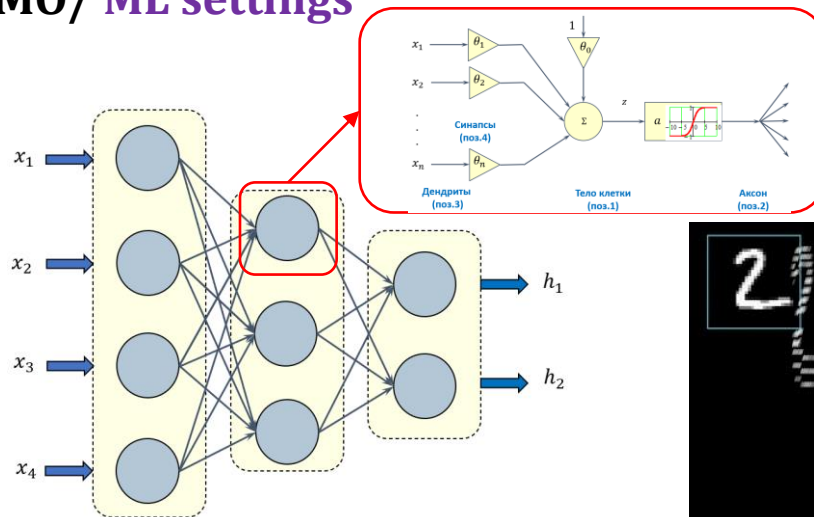
10^{-16} Дж

10^{-6} Дж



Лекция 2. Искусственные нейронные сети (ИНС) и обучение с учителем / Lecture 2. Artificial Neural Networks (ANNs) and Supervised Learning

1. Напоминание / Contents of the previous lecture
2. Естественные и искусственные нейронные сети (ИНС)/ Natural and artificial neural networks (ANNs)
3. Прямые вычисления в ИНС / Forward propagation
4. Обратные вычисления в ИНС/ Back propagation
5. Настройка МО/ ML settings



Прямые вычисления в ИНС / Forward propagation

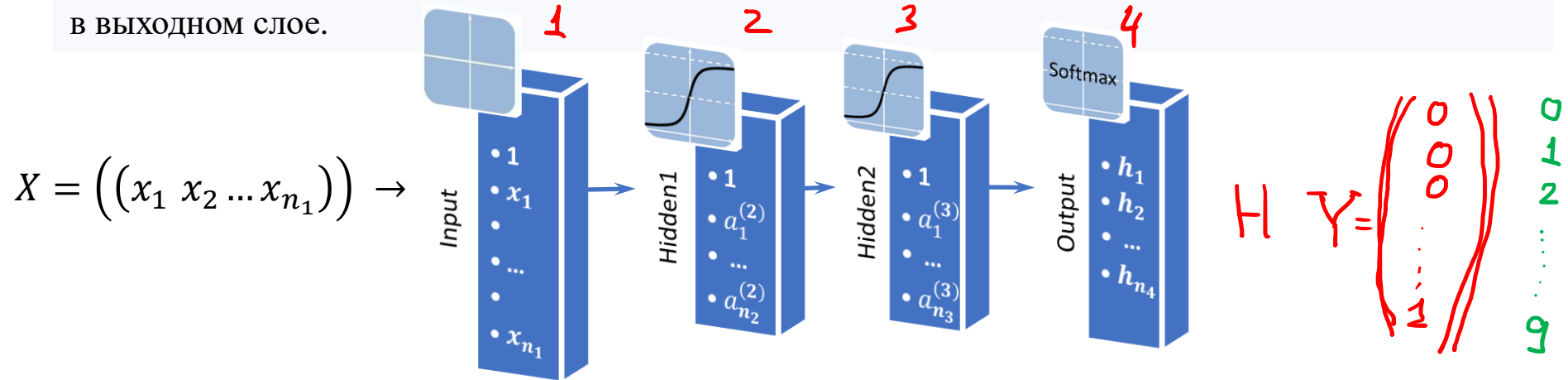
Обучение с учителем подразумевает наличие правильных ответов (**labeled data**), которые можно сравнить с результатами вычислений ИНС.

Задача распознавания (классификации) рукописных чисел

Архитектура ИНС: количество слоев - l ; количество нейронов в k -ом слое - n_k (n_l - количество классов); логистическая функция активации в скрытых слоях и функция активации «софтмакс» в выходном слое.



Gray scale picture of "Nine"



Вычисления в прямом направлении ИНС, расчет матриц $A^{(k)}$ результатов в каждом слое / **Forward propagation**

Сл.1 (входной). На вход слоя подается дополненная единицей матрица X . На выходе то же: $A^{(1)} = \begin{pmatrix} 1 & X \end{pmatrix}$.

Сл.2 (скрытый). Данные с 1^{го} слоя умнож. на веса $\Theta^{(1)}$ и сумм.: $Z^{(2)} = A^{(1)}\Theta^{(1)}$. Затем прим. ф-я актив.: $A^{(2)} = \begin{pmatrix} 1 & \text{sigmoid}(Z^{(2)}) \end{pmatrix}$.

Сл.3 (скрытый). $Z^{(3)} = A^{(2)}\Theta^{(2)}$, $A^{(3)} = \begin{pmatrix} 1 & \text{sigmoid}(Z^{(3)}) \end{pmatrix}$.

Сл.4 (выходной). $Z^{(4)} = A^{(3)}\Theta^{(3)}$, $A^{(4)} = \text{softmax}(Z^{(4)}) = H$.

Прямые вычисления в ИНС / Forward propagation

Напоминание. Вычисления в задаче логистической регрессии и бинарной классификации:

$$X = \begin{pmatrix} \begin{pmatrix} x_1^{(1)} & x_2^{(1)} \\ \dots & \dots \\ x_1^{(m)} & x_2^{(m)} \end{pmatrix} \end{pmatrix}; Y = \begin{pmatrix} \begin{pmatrix} y^{(1)} \\ \dots \\ y^{(m)} \end{pmatrix} \end{pmatrix} \rightarrow X = \begin{pmatrix} \begin{pmatrix} 1 & x_1^{(1)} & x_2^{(1)} \\ \dots & \dots & \dots \\ 1 & x_1^{(m)} & x_2^{(m)} \end{pmatrix} \end{pmatrix}; \Theta = \begin{pmatrix} \begin{pmatrix} \theta_0 \\ \theta_1 \\ \theta_2 \end{pmatrix} \end{pmatrix}; Z = X\Theta; \rightarrow h(z) = \frac{1}{1+e^{-z}}.$$

Вычисления в ИНС. Количество слоев - l ; количество нейронов в k -ом слое - n_k (n_l - количество классов); логистическая функция активации в скрытых слоях и функция активации «софтмакс» в выходном слое.

Сл.1 (входной). На вход слоя подается дополненная единицей матрица X . На выходе то же: $A^{(1)} = ((1 \ X))$.

Сл.2 (скрытый). Данные с 1^{го} слоя умнож. на веса $\Theta^{(1)}$ и сумм.: $Z^{(2)} = A^{(1)}\Theta^{(1)}$. Затем прим. ф-я актив.: $A^{(2)} = ((1 \ \text{sigmoid}(Z^{(2)})))$.

Сл.3 (скрытый). $Z^{(3)} = A^{(2)}\Theta^{(2)}$, $A^{(3)} = ((1 \ \text{sigmoid}(Z^{(3)})))$.

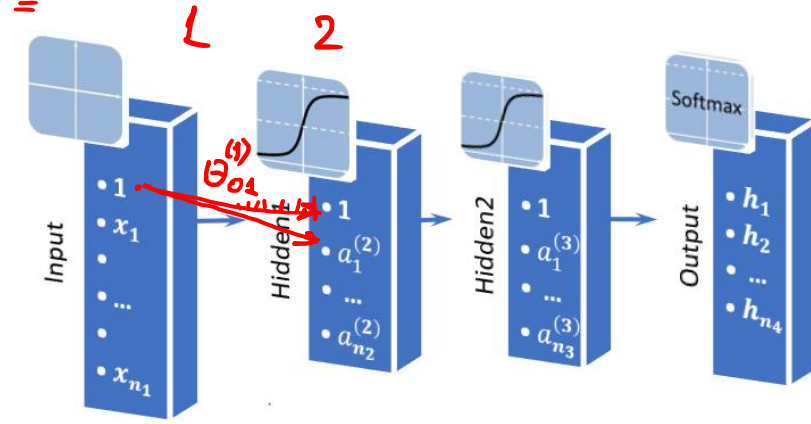
Сл.4 (выходной). $Z^{(4)} = A^{(3)}\Theta^{(3)}$, $A^{(4)} = \text{softmax}(Z^{(4)}) = H$.

$$X, Y; \rightarrow A^{(1)} = ((1 \ x_1 \dots x_{n_1})), \Theta^{(1)} = \begin{pmatrix} \begin{pmatrix} \theta_{01}^{(1)} & \theta_{02}^{(1)} & \theta_{0n_2}^{(1)} \\ \theta_{11}^{(1)} & \theta_{12}^{(1)} & \dots & \theta_{1n_2}^{(1)} \\ \dots & \dots & \dots & \dots \\ \theta_{n_11}^{(1)} & \theta_{n_12}^{(1)} & \dots & \theta_{n_1n_2}^{(1)} \end{pmatrix} \end{pmatrix}; Z^{(2)} = A^{(1)}\Theta^{(1)} \text{ или } z_j^{(2)} = \underline{a_i^{(1)}} \underline{\theta_{ij}^{(1)}}; \rightarrow$$

$$\rightarrow A^{(2)} = ((1 \ \text{sigmoid}(Z^{(2)}))) \text{ или } a_0^{(2)} = 1, a_j^{(2)} = \frac{1}{1+e^{-z_j^{(2)}}}, (j = 1, \dots, n_2).$$

...

$$\rightarrow A^{(4)} = \text{softmax}(Z^{(4)}) \text{ или } a_j^{(4)} = \frac{e^{z_j^{(4)}}}{\sum_{i=1}^{(n_4+1)} e^{z_i^{(4)}}}.$$

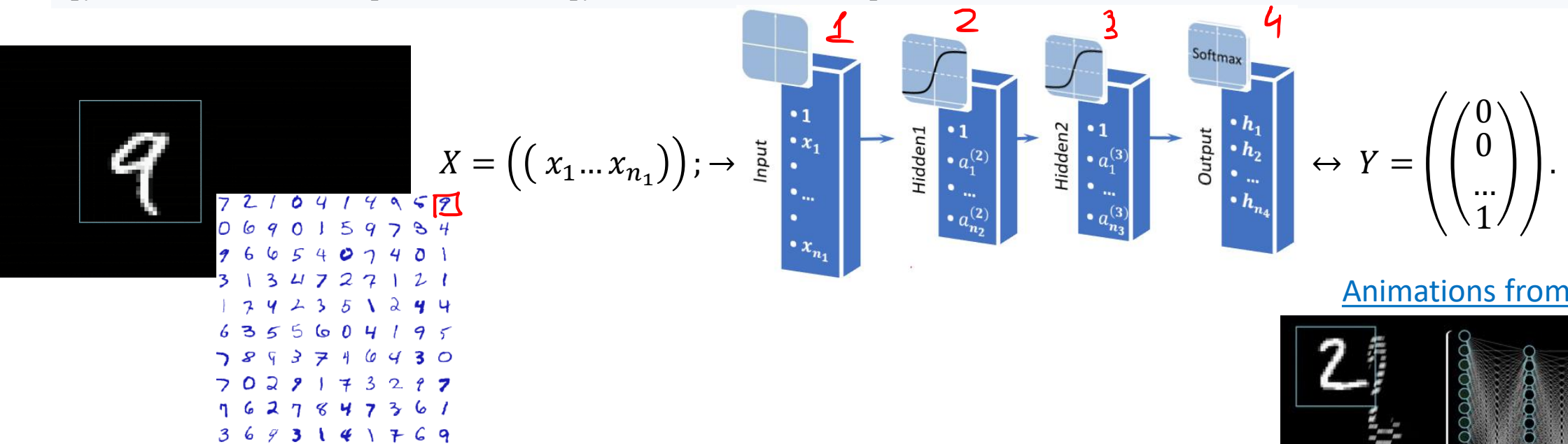


Прямые вычисления в ИНС / Forward propagation

Напоминание. Функция качества в задаче логистической регрессии.

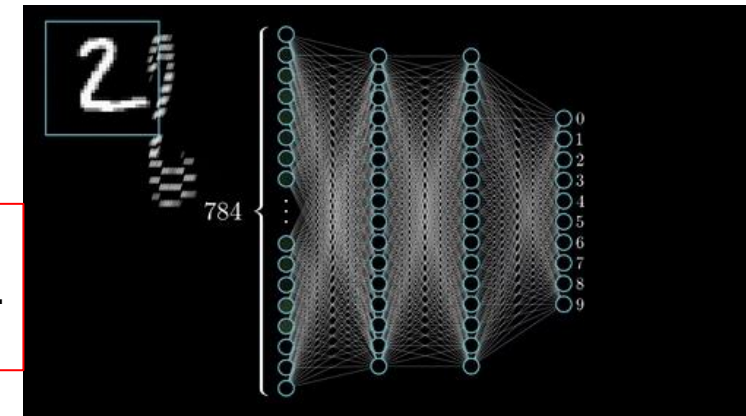
$$J(\Theta) = -\frac{1}{m} \sum_{i=1}^m (y^{(i)} \ln(h^{(i)}) + (1 - y^{(i)}) \ln(1 - h^{(i)})) + \frac{\lambda}{2m} \sum_{j=1}^n \theta_j^2 \Rightarrow \min.$$

Функция качества в ИНС. Количество слоев - l ; количество нейронов в k -ом слое - n_k (n_l - количество классов); логистическая функция активации в скрытых слоях и функция активации «софтмакс» в выходном слое.



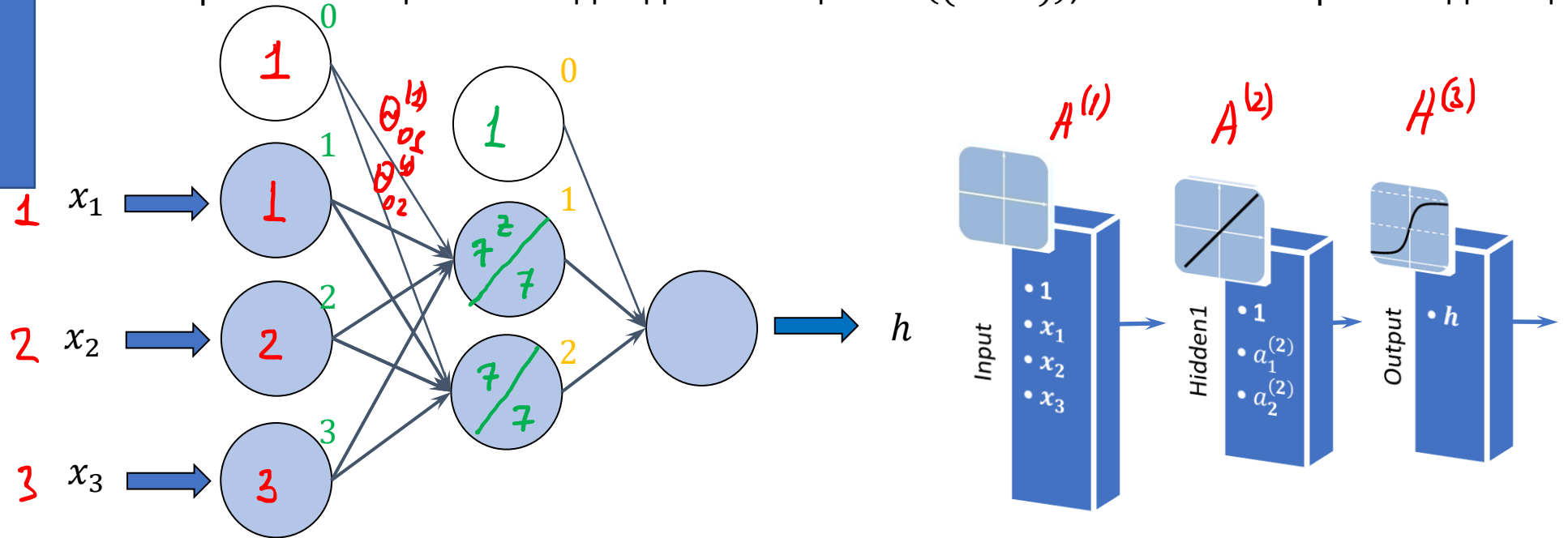
[Animations from "3Blue1Brown"](#)

$$J(\Theta^{(k)}) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{n_l} (y_j^{(i)} \ln(h_j^{(i)}) + (1 - y_j^{(i)}) \ln(1 - h_j^{(i)})) + \frac{\lambda}{2m} \sum_{k=1}^{l-1} \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k+1}} (\theta_{ij}^{(k)})^2 \Rightarrow \min.$$



Прямые вычисления в ИНС / Forward propagation

Пример. ИНС содержит 3 вх. нейр., 2 нейр. в скр. слое с лин. функ. активации и 1 вых. нейрон с логист. ф-ей активации. На вход подается м-ца $X = ((1 \ 2 \ 3))$, все веса сети равны единице.



Прямые вычисления в ИНС с количеством нейронов $N = [3 \ 2 \ 1]$, все синаптические веса равны единице $\theta_{ij}^{(k)} = 1$.

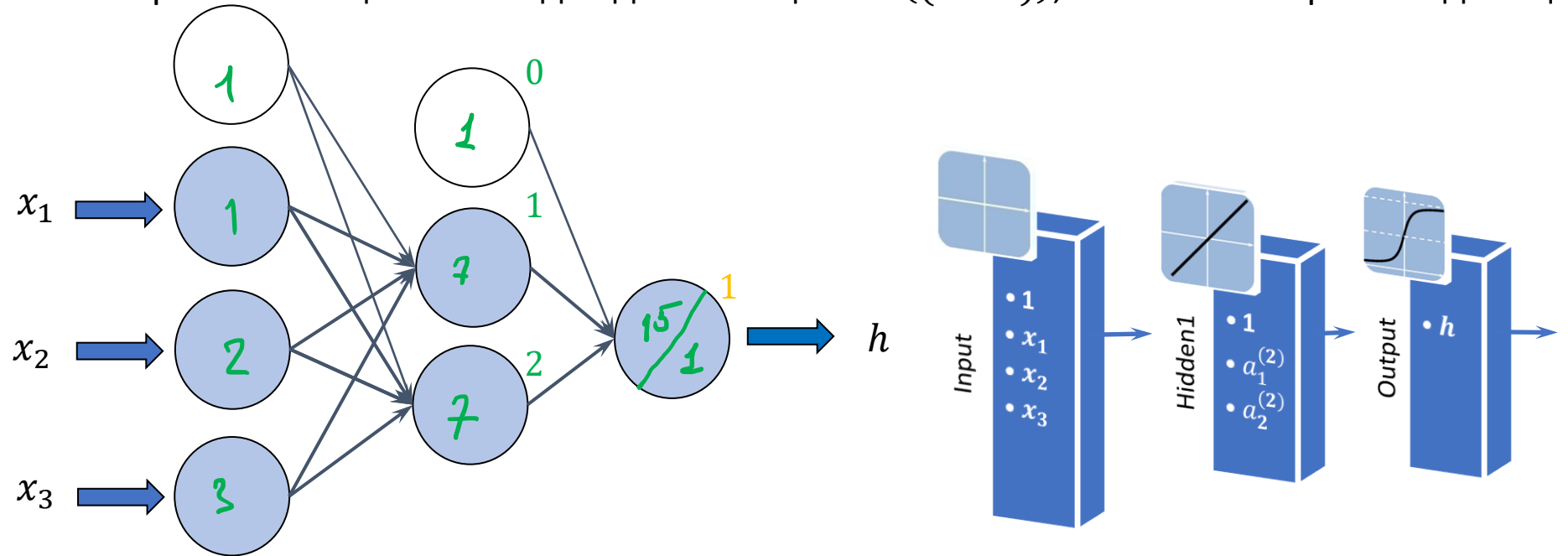
$$X = ((1 \ 2 \ 3)); \rightarrow A^{(1)} = ((1 \ 1 \ 2 \ 3)), \Theta^{(1)} = \begin{pmatrix} \theta_{01}^{(1)} & \theta_{02}^{(1)} \\ \theta_{11}^{(1)} & \theta_{12}^{(1)} \\ \theta_{21}^{(1)} & \theta_{22}^{(1)} \\ \theta_{31}^{(1)} & \theta_{32}^{(1)} \end{pmatrix}; \quad Z^{(2)} = A^{(1)}\Theta^{(1)} \text{ или } z_j^{(2)} = a_i^{(1)}\theta_{ij}^{(1)}; \quad z_1^{(2)} = a_0^{(1)}\theta_{01}^{(1)} + a_1^{(1)}\theta_{11}^{(1)} + a_2^{(1)}\theta_{21}^{(1)} + a_3^{(1)}\theta_{31}^{(1)}$$

$\rightarrow A^{(2)} = ((1 \ Z^{(2)}))$ или $a_0^{(2)} = 1, a_j^{(2)} = z_j^{(2)}, \text{size}(\Theta^{(2)}) = [3 \ 1]; \rightarrow$

$\theta_{ij}^{(k)}$ - # нейр. "k+1" слоя
нейрона k-го слоя

Прямые вычисления в ИНС / Forward propagation

Пример. ИНС содержит 3 вх. нейр., 2 нейр. в скр. слое с лин. функ. активации и 1 вых. нейрон с логист. ф-ей активации. На вход подается м-ца $X = ((1 \ 2 \ 3))$, все веса сети равны единице.



Прямые вычисления в ИНС с количеством нейронов $N = [3 \ 2 \ 1]$, все синаптические веса равны единице $\theta_{ij}^{(k)} = 1$.

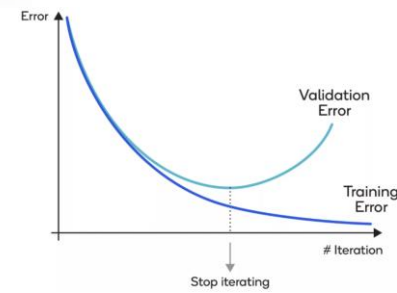
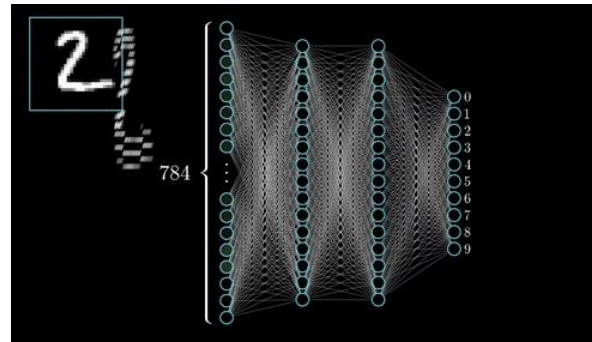
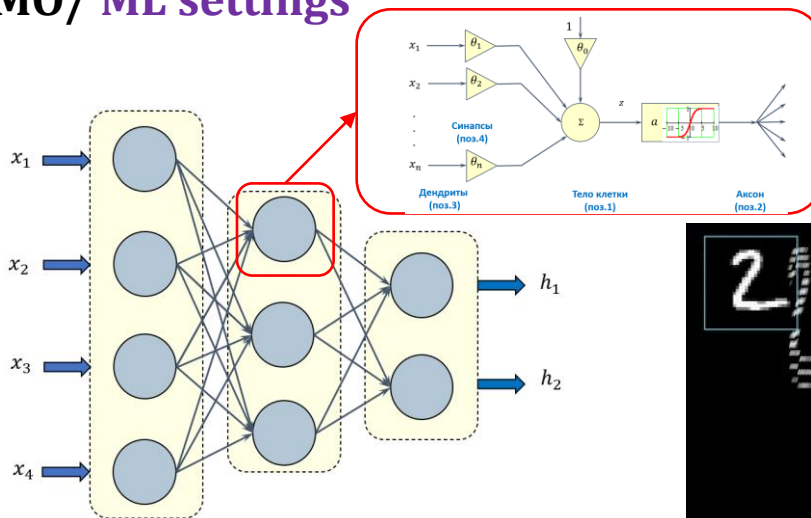
$$\rightarrow A^{(2)} = ((1 \ 7 \ 7)), \Theta^{(2)} = \begin{pmatrix} \theta_{01}^{(2)} \\ \theta_{11}^{(2)} \\ \theta_{21}^{(2)} \end{pmatrix}; Z^{(3)} = A^{(2)} \Theta^{(2)} \text{ или } z_1^{(3)} = a_i^{(2)} \theta_{i1}^{(2)}; z_1^{(3)} = a_0^{(2)} \theta_{01}^{(2)} + a_1^{(2)} \theta_{11}^{(2)} + a_2^{(2)} \theta_{21}^{(2)} \rightarrow$$

$$\rightarrow A^{(3)} = h = \text{sigmoid}(Z^{(3)}) \text{ или } a_1^{(3)} = h = \frac{1}{1 + e^{-z_1^{(3)}}} \approx 1$$

слой ИНС
нейр. "k+1" слоя
нейрона k-го слоя

Лекция 2. Искусственные нейронные сети (ИНС) и обучение с учителем / Lecture 2. Artificial Neural Networks (ANNs) and Supervised Learning

1. Напоминание / Contents of the previous lecture
2. Естественные и искусственные нейронные сети (ИНС)/ Natural and artificial neural networks (ANNs)
3. Прямые вычисления в ИНС / Forward propagation
4. Обратные вычисления в ИНС/ Back propagation
5. Настройка МО/ ML settings



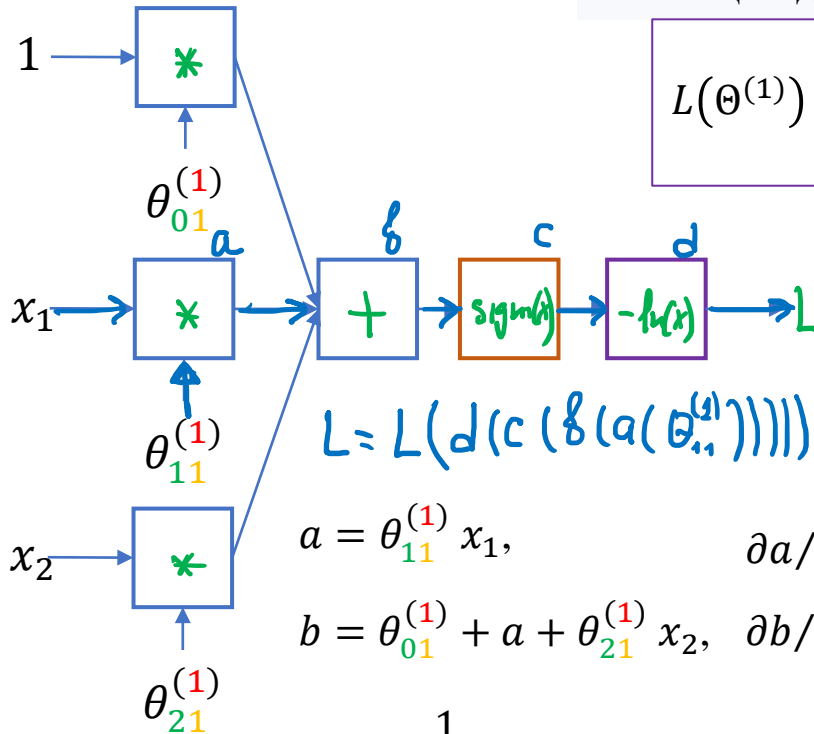
Обратный расчет и обучение ИНС / Backpropagation

Обратный расчет выполняется с целью определения компонент градиента функции качества и является этапом процесса обучения:

$$L(\Theta^{(k)}), \rightarrow \left[\frac{\partial L}{\partial \theta_{ij}^{(k)}} \right], \rightarrow \theta_{ij}^{(k)} \rightarrow L(\Theta^{(k)}), \rightarrow \dots$$

Пример. ИНС с 2 входными нейронами и 1 выходным нейроном с логистической функцией активации для бинарной классификации. Функция качества для дата сета из m образцов имеет вид:

$$L(\Theta^{(1)}) = - \sum_{i=1}^m (y^{(i)} \ln(h^{(i)})) = - \sum_{i=1}^m \ln(h^{(i)}) \Rightarrow \min.$$



$$L = L(d(c(b(a(\theta_{01}^{(1)}))))))$$

$$a = \theta_{11}^{(1)} x_1,$$

$$\partial a / \partial \theta_{11}^{(1)} = x_1,$$

$$b = \theta_{01}^{(1)} + a + \theta_{21}^{(1)} x_2, \quad \partial b / \partial a = 1,$$

$$c = \frac{1}{1 + e^{-b}},$$

$$\partial c / \partial b = c(1 - c),$$

$$d = -\ln(c),$$

$$\partial d / \partial c = -1/c,$$

$$d = L.$$

$$\partial L / \partial d = 1.$$

Пусто $m=1$; найти $\frac{\partial L}{\partial \theta_{11}^{(1)}}$ $X = ((x_1 \ x_2))$

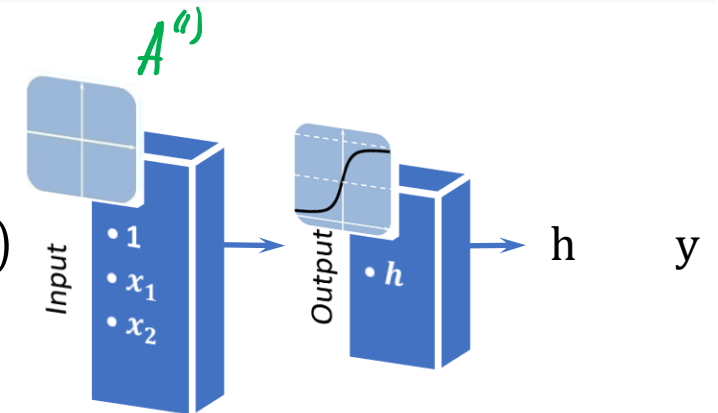
$$\frac{\partial L}{\partial \theta_{11}^{(1)}} = \frac{\partial L}{\partial d} \cdot \frac{\partial d}{\partial c} \cdot \frac{\partial c}{\partial b} \cdot \frac{\partial b}{\partial a} \cdot \frac{\partial a}{\partial \theta_{11}^{(1)}} =$$

$$= 1 \left(-\frac{1}{c} \right) c(1-c) 1 x_1 = (c-1) x_1.$$

Пусто $m>1$

$$\frac{\partial L}{\partial \theta_{11}^{(1)}} = \sum_{i=1}^m (h^{(i)} - 1) x_1^{(i)},$$

$$\frac{\partial L}{\partial \theta_{21}^{(1)}} = \dots \quad \frac{\partial L}{\partial \theta_{01}^{(1)}} = \dots$$



$$1.0. \quad f(u) = (10x - 5)^2 = u^2$$

$$u = 10x - 5; \quad f(u(x))$$

$$\frac{df}{dx} = \frac{df}{du} \cdot \frac{du}{dx} =$$

$$= 2(10x - 5) \cdot 10 =$$



Обратный расчет и обучение ИНС / Backpropagation

Обратный расчет выполняется с целью определения компонент градиента функции качества и является этапом процесса обучения:

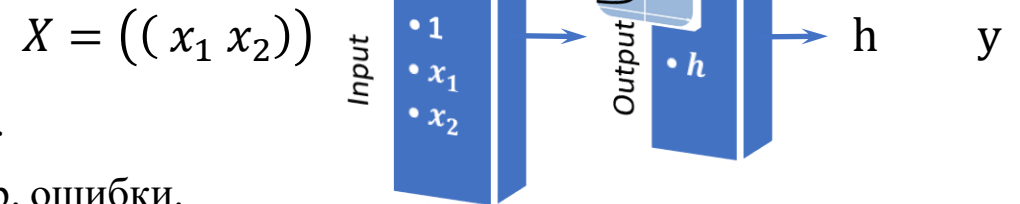
$$L(\Theta^{(k)}), \rightarrow \left[\left[\partial L / \partial \theta_{ij}^{(k)} \right] \right], \rightarrow \theta_{ij}^{(k)} \rightarrow L(\Theta^{(k)}), \rightarrow \dots$$

Пример. ИНС с 2 входными нейронами и 1 выходным нейроном с логистической функцией активации для бинарной классификации. Функция качества для дата сета из m образцов имеет вид:

$$L(\Theta^{(1)}) = - \sum_{i=1}^m \left(y^{(i)} \ln(h^{(i)}) \right) = - \sum_{i=1}^m \ln(h^{(i)}) \Rightarrow \min.$$

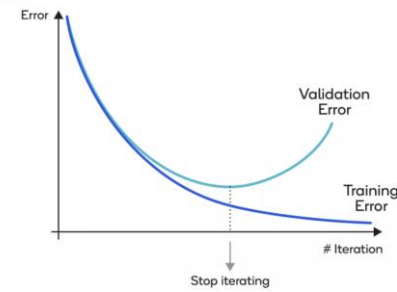
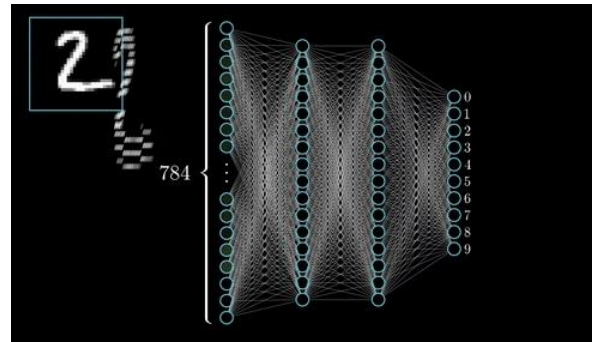
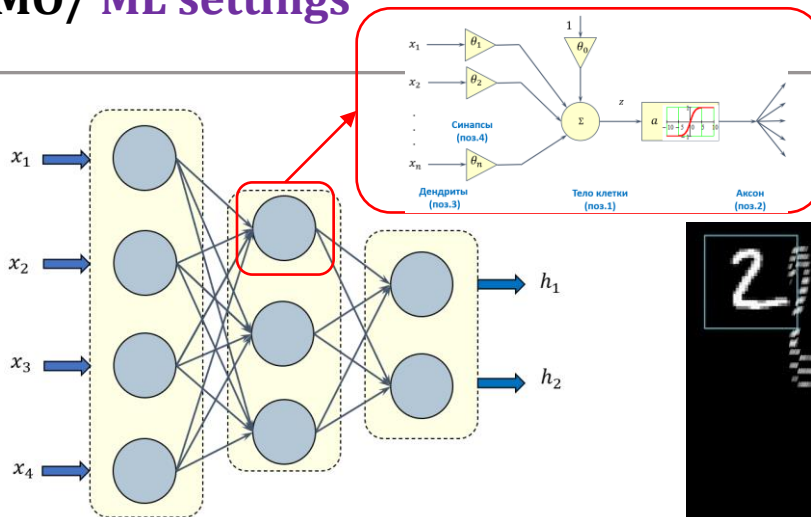
Алгоритм обучения (обобщенный).

1. Задать начальные значения компонент матрицы $\Theta^{(k)}$ случайным образом.
2. Рассчитать вектор градиента $\nabla L = \left[\left[\partial L / \partial \theta_{ij}^{(k)} \right] \right]$ методом обратного распр. ошибки.
3. Найти новые значения компонент Θ : $\theta_{ij}^{(k)H} = \theta_{ij}^{(k)C} - \alpha \frac{\partial L}{\partial \theta_{ij}^{(k)}}$.
4. Повторять пп. 2-3 до достижения минимума L : $L^H - L^C < \delta$ или #итерации $> N_{max}$.
5. Вывод результатов: $\Theta^{(k)}$.



Лекция 2. Искусственные нейронные сети (ИНС) и обучение с учителем / Lecture 2. Artificial Neural Networks (ANNs) and Supervised Learning

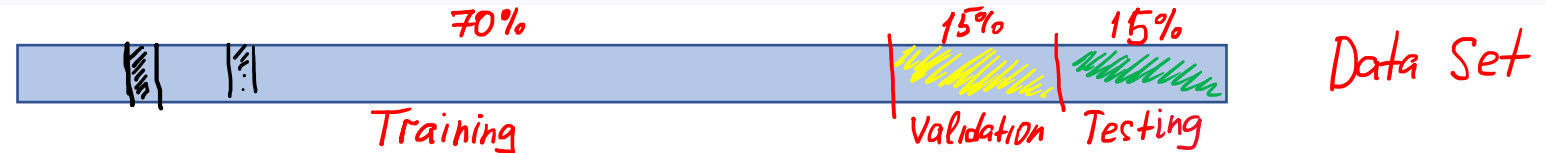
1. Напоминание / Contents of the previous lecture
2. Естественные и искусственные нейронные сети (ИНС)/ Natural and artificial neural networks (ANNs)
3. Прямые вычисления в ИНС / Forward propagation
4. Обратные вычисления в ИНС/ Back propagation
5. Настройка МО/ ML settings



Настройка моделей MO/ ML settings

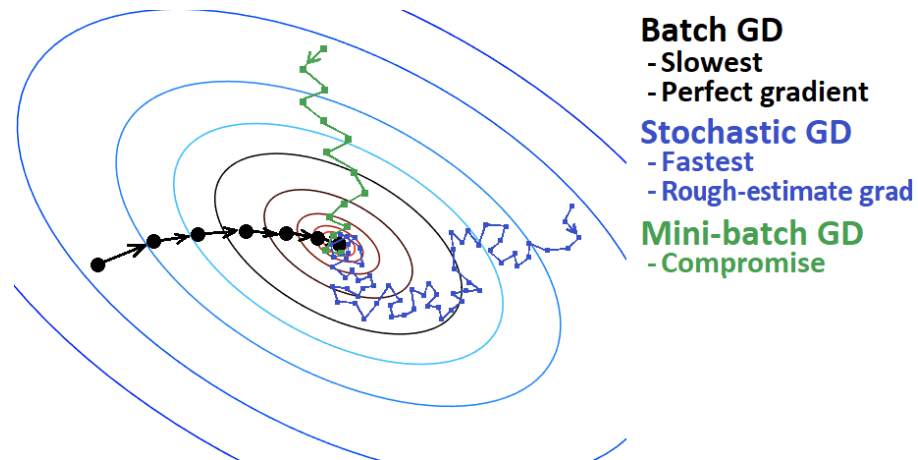
Параметры модели определяются в ходе решения задачи МО.

Гиперпараметры задаются пользователем, как правило не единственным образом, и их значения влияют на значения искомых параметров.



$$J(\Theta^{(k)}) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{n_l} (y_j^{(i)} \ln(h_j^{(i)}) + (1 - y_j^{(i)}) (\ln(1 - h_j^{(i)}))) + \frac{\lambda}{2m} \sum_{k=1}^{l-1} \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k+1}} (\theta_{ij}^{(k)})^2 \Rightarrow \min.$$

1. Масштабирование признаков / Feature Scaling
2. Скорость обучения α / Learning rate
3. Погрешность δ и количество итераций N_{\max} / Error and # of iterations
4. Количество данных для градиентного метода / Batch gradient descent (GD) – Mini-Batch GD. – Stochastic GD
5. Регуляризация / Regularization



Настройка моделей МО/ ML settings

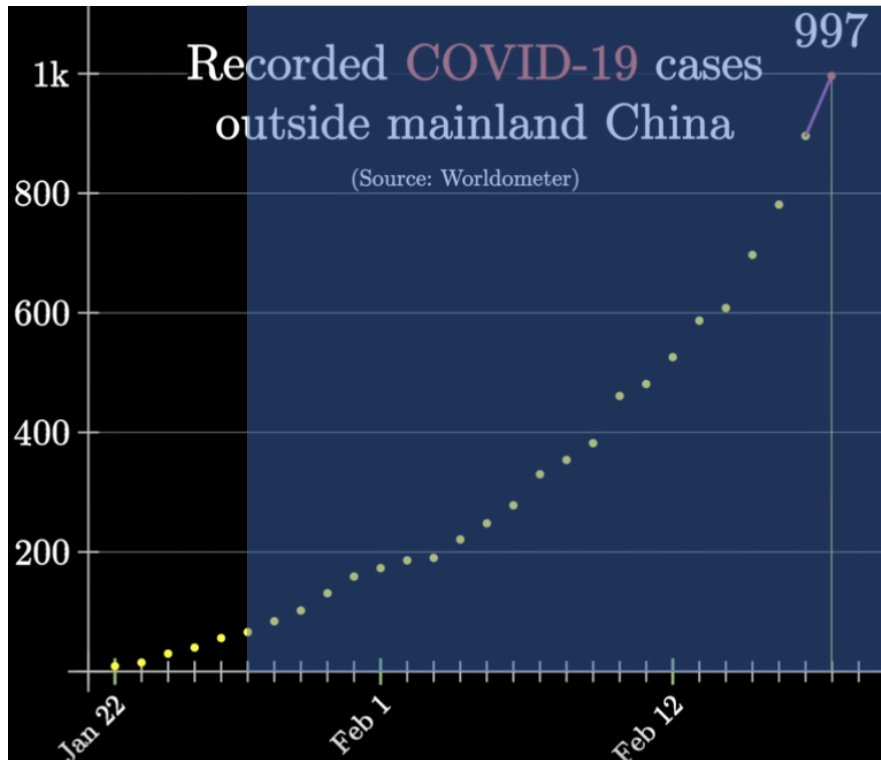
Параметры модели определяются в ходе решения задачи МО.

Гиперпараметры задаются пользователем, как правило не единственным образом, и их значения **влияют** на значения искомых параметров.

...

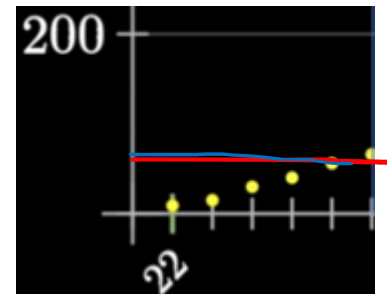
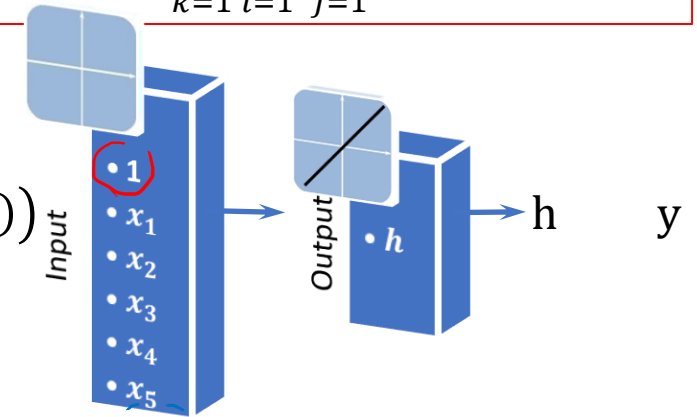
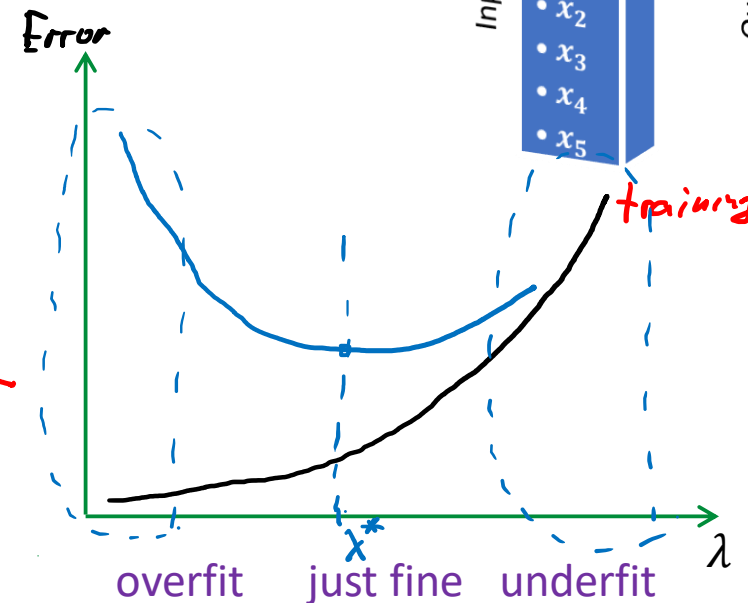
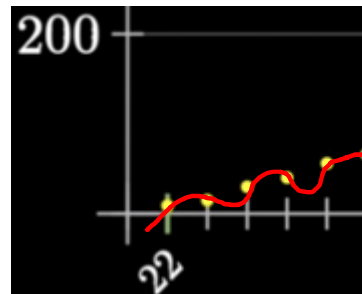
5. Регуляризация / Regularization

$$J(\Theta^{(k)}) = -\frac{1}{m} \sum_{i=1}^m \sum_{j=1}^{n_l} (y_j^{(i)} \ln(h_j^{(i)}) + (1 - y_j^{(i)}) (\ln(1 - h_j^{(i)}))) + \frac{\lambda}{2m} \sum_{k=1}^{l-1} \sum_{i=1}^{n_k} \sum_{j=1}^{n_{k+1}} (\theta_{ij}^{(k)})^2 \Rightarrow \min.$$



Exponential Growth and Epidemics

$$\begin{aligned} x_1 &= x; \\ x_1 &= x^2; \\ &\dots \\ x_5 &= x^5; \end{aligned}$$



Самостоятельная работа / Homework

Вопросы и задания.

1. Изобразите архитектуры простейших нейронных сетей, вычисления в которых идентичны вычислениям при линейной и логистической регрессии.
2. Каким образом в ИНС хранятся знания и как они из ИНС извлекаются?
3. Позволяет ли применение метода **Mini-Batch GD** решить проблему поиска глобального экстремума при наличии локальных?