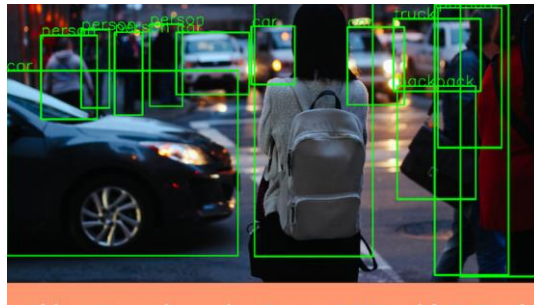
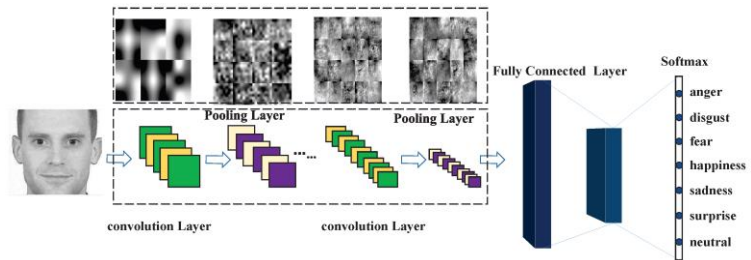


Основы искусственного интеллекта



Лекция 7 Введение в ИНС

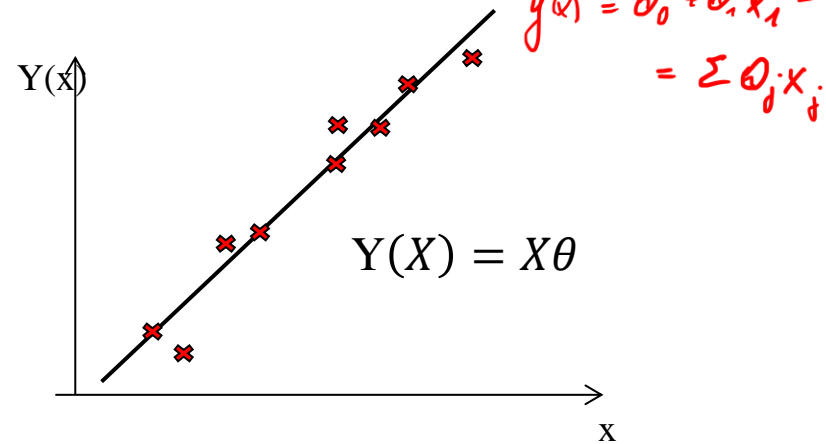
к.ф.-м.н., доцент кафедры ИСиЦТ
Корнаева Е.П.

Понятие ИНС

От простого к сложному

Задача регрессии

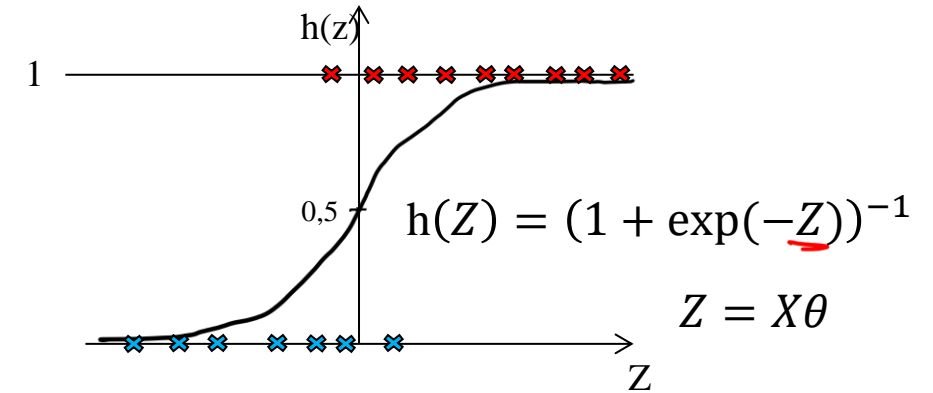
$$X \in R, Y \in R$$



Y – непрерывная величина

Задача классификации

$$Y \in \{0; 1\}, X \in R$$



Y – дискретная величина

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{bmatrix}$$

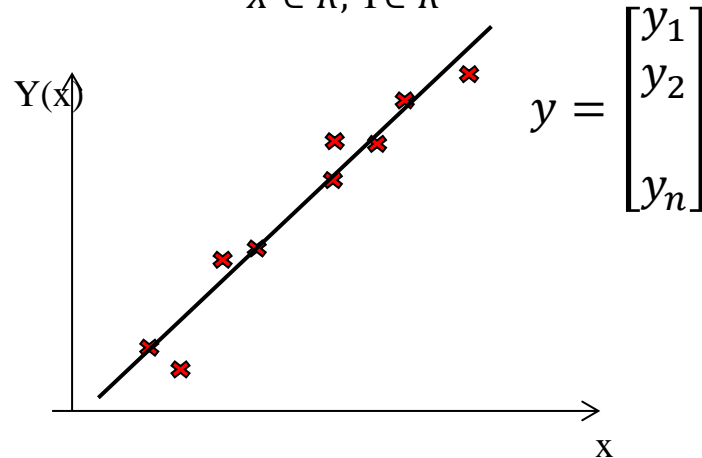
$$y = \begin{bmatrix} 1 \\ 0 \\ 0 \\ \dots \\ 1 \end{bmatrix}$$

Понятие ИНС

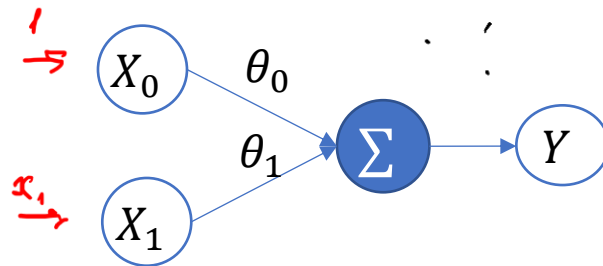
От простого к сложному

Задача аппроксимации

$$X \in R, Y \in R$$

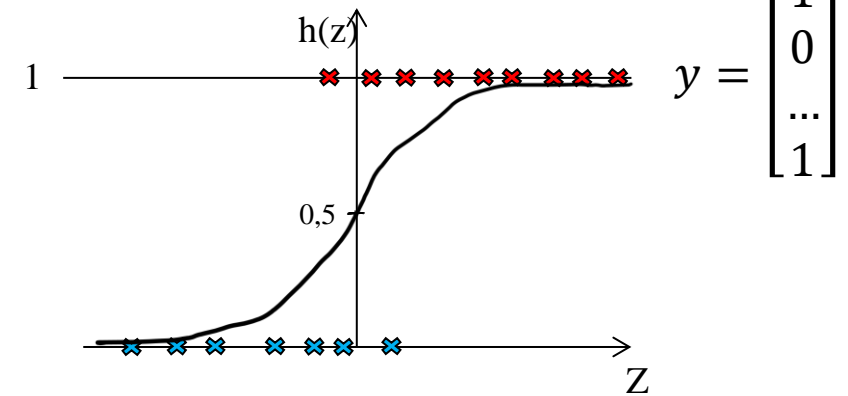


$$Y(X) = X\theta$$

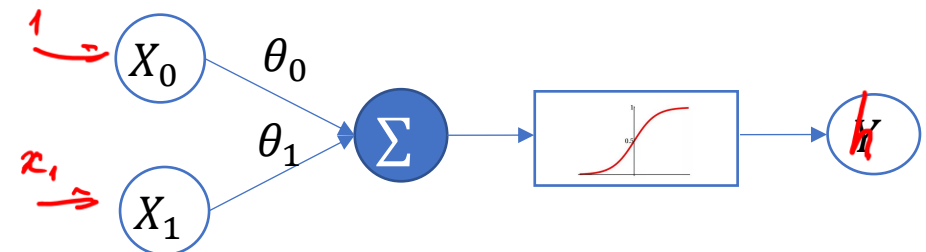


Задача классификации

$$Y \in \{0; 1\}, X \in R$$



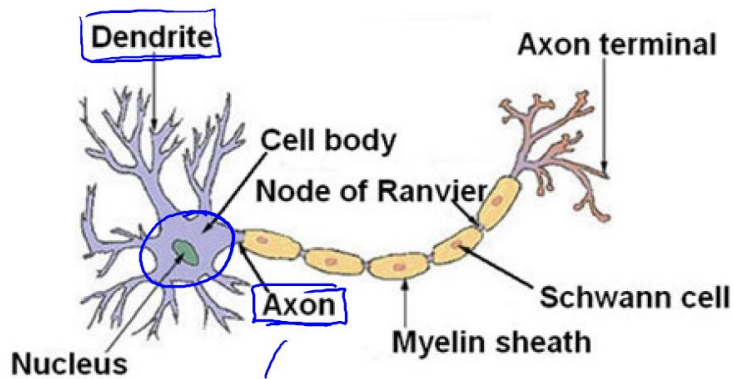
$$Y(Z) = (1 + \exp(-X\theta))^{-1}$$



Введение в ИНС

Понятие ИНС. Биологическая аналогия

ИНС - математическая модель, а также её программное или аппаратное воплощение, построенная по принципу организации и функционирования биологических нейронных сетей.



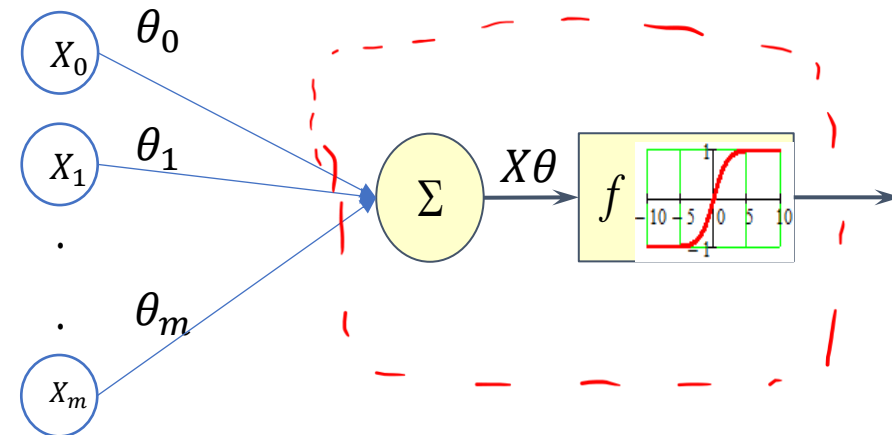
дендриты – получают информацию;

тело клетки – обрабатывает информацию;

аксон – передает обработанную информацию другим нейронам;

синапсы – соединяют аксон и дендриты других нейронов.

Однослойный перцептрон



	X_0	X_1	...	X_m
1	1	x_{11}		x_{1m}
2	1	x_{21}		x_{2m}
...			...	
n	1	x_{n1}		x_{nm}

X

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \\ \dots \\ \theta_m \end{bmatrix}$$

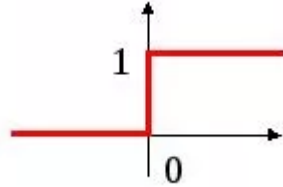
Введение в ИНС

Понятие ИНС

Функции активации

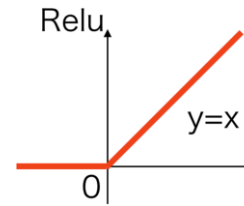
Ступенчатая функция

$$f(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

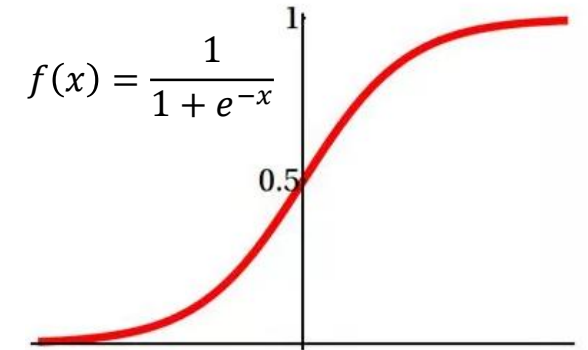


ReLU

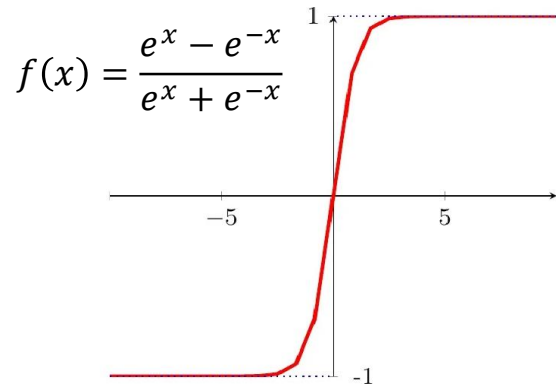
$$f(x) = \begin{cases} x, & x \geq 0 \\ 0, & x < 0 \end{cases}$$



Сигмоида

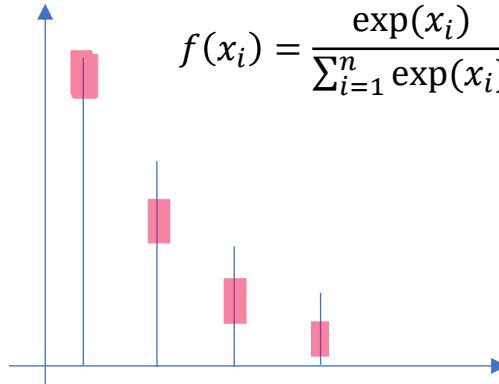


Гиперболический тангенс



softmax

$$f(x_i) = \frac{\exp(x_i)}{\sum_{i=1}^n \exp(x_i)}$$



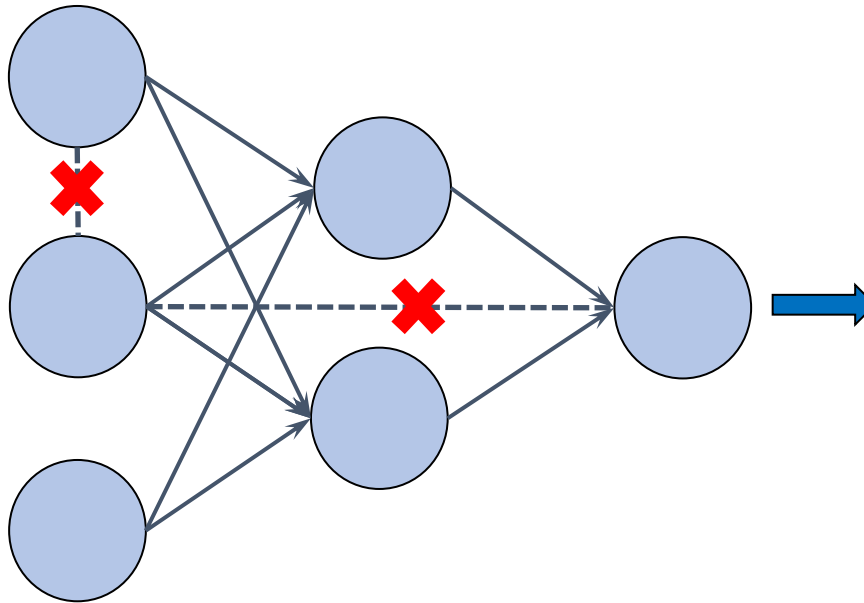
...

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

Нейроны слоя не связаны.

Нейроны передают информацию только нейронам следующего слоя.

Перепрыгивание через слои запрещено.



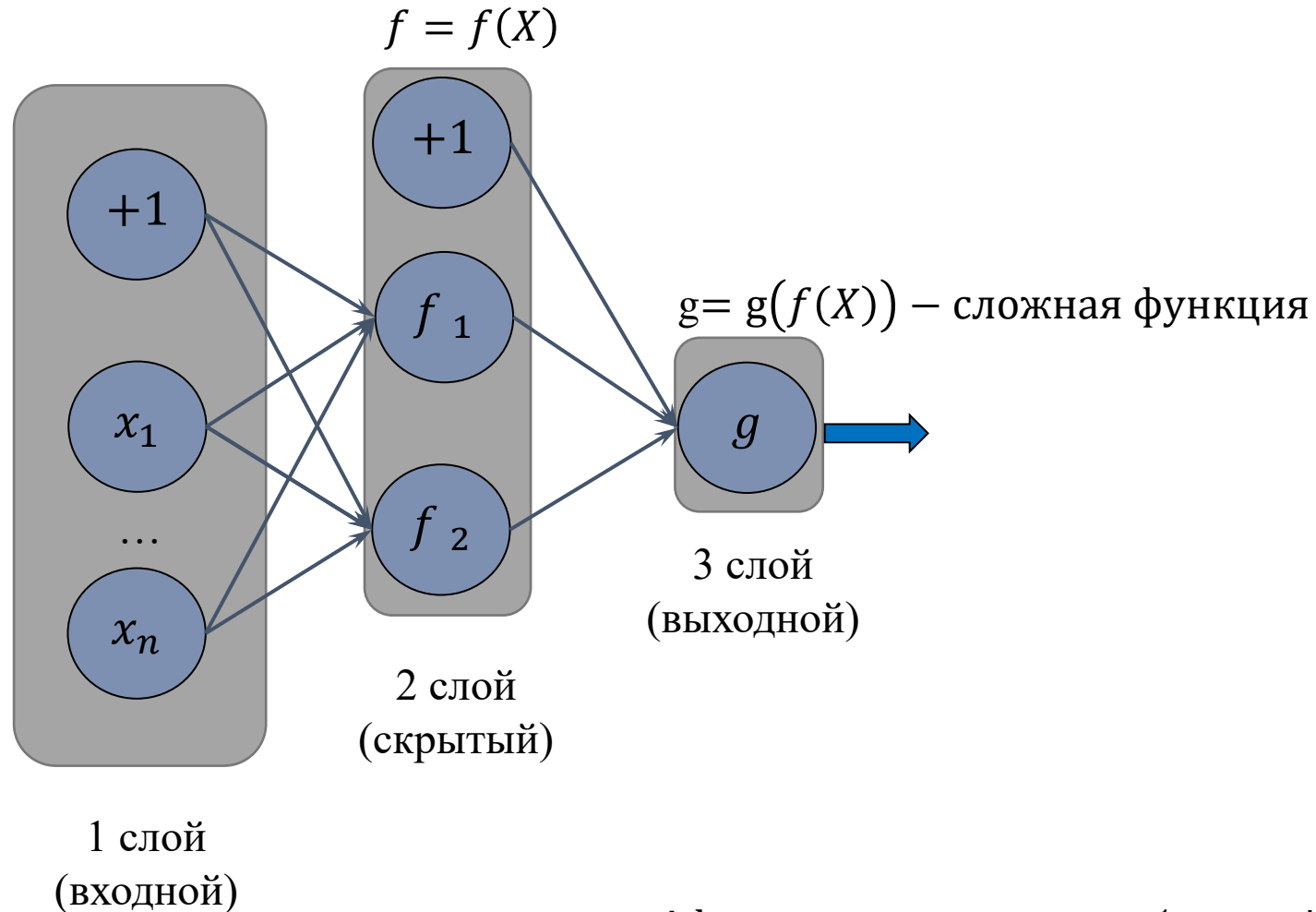
Направление потока информации



Архитектура сетей прямого распространения / Feed forward neural network

Пример.

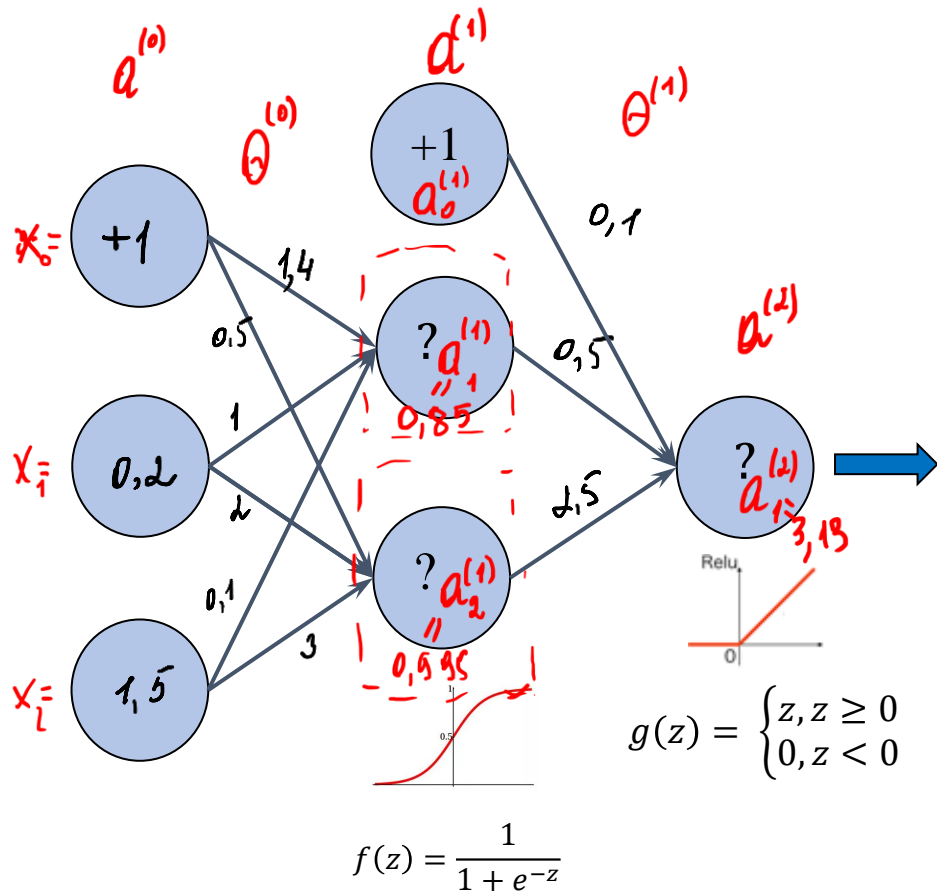
3х – слойная ИНС прямого распространения



! функция активации м. б. разной на разных слоях

Архитектура сетей прямого распространения / Feed forward neural network

Пример. Вычислить значение функции активации $g(z)$ на последнем слое.

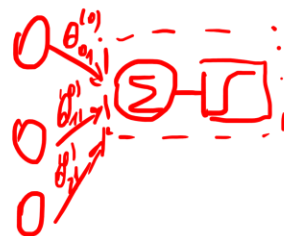


$$a_1^{(1)} = \text{sigm} \left(\sum x \theta^{(0)} \right) = \frac{1}{1 + \exp(-\theta_0 x_0 - \theta_1 x_1 - \theta_2 x_2)} =$$

$$\approx \frac{1}{1 + \exp(-1.4 \cdot 1 - 1 \cdot 0.2 - 0.1 \cdot 1.5)} \approx 0.85$$

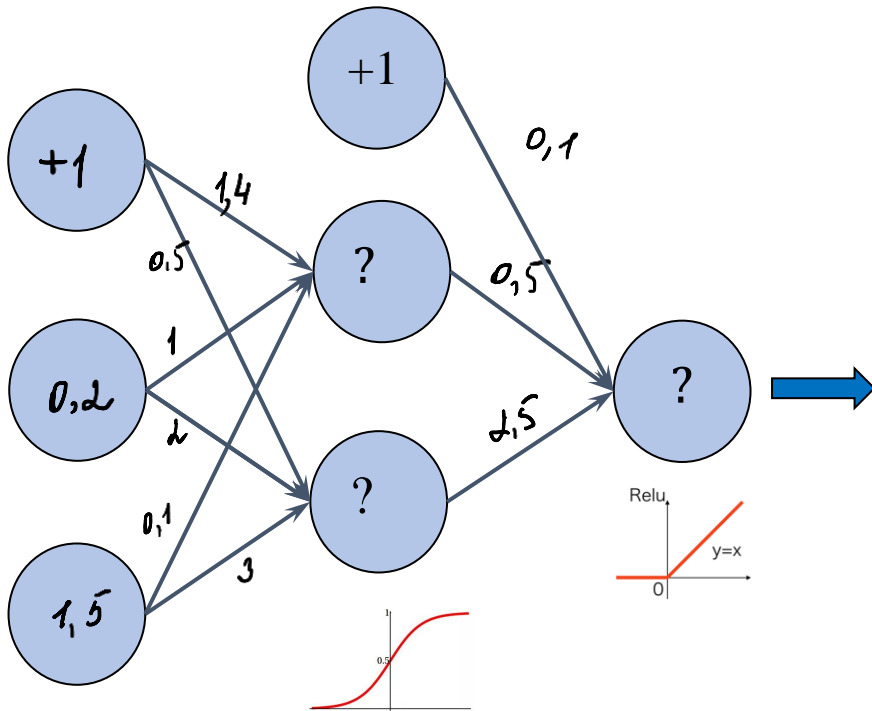
$$a_2^{(1)} = \text{sigm} \left(\sum x \theta^{(0)} \right) = \left(1 + \exp(-0.5 \cdot 1 - 2 \cdot 0.2 - 3 \cdot 1.5) \right)^{-1} \approx 0.995$$

$$a_1^{(2)} = \text{Relu} \left(\sum x \theta^{(1)} \right) = \text{Relu} \left(0.1 \cdot 1 + 0.5 \cdot 0.85 + 2.5 \cdot 0.995 \right) = 3.13$$



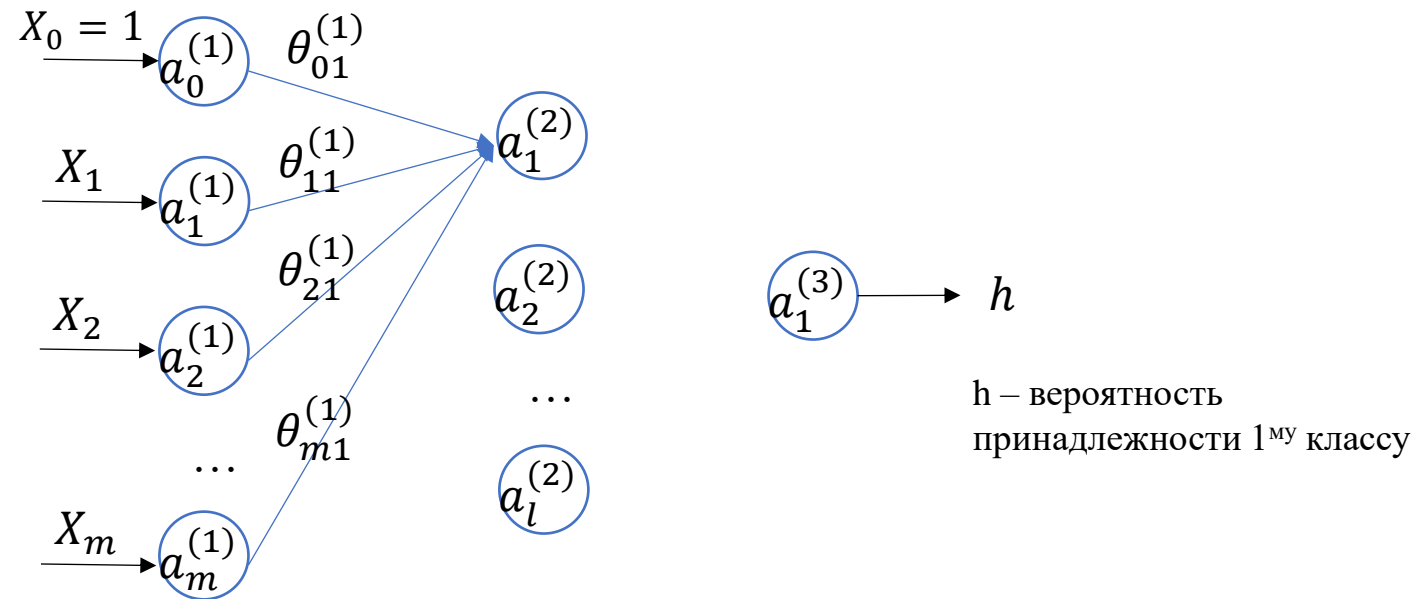
Архитектура сетей прямого распространения / Feed forward neural network

Пример. Запишите вычисления в матричном виде.



Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: $Y=\{0;1\}$



$\theta_{pq}^{(1)}$ — весовые коэффициенты 1 слоя

p — номер нейрона предыдущего слоя;

q — номер нейрона текущего слоя;

$$a_1^{(2)} = g(\theta_{01}^{(1)} a_0^{(1)} + \theta_{11}^{(1)} a_1^{(1)} + \dots + \theta_{m1}^{(1)} a_m^{(1)})$$

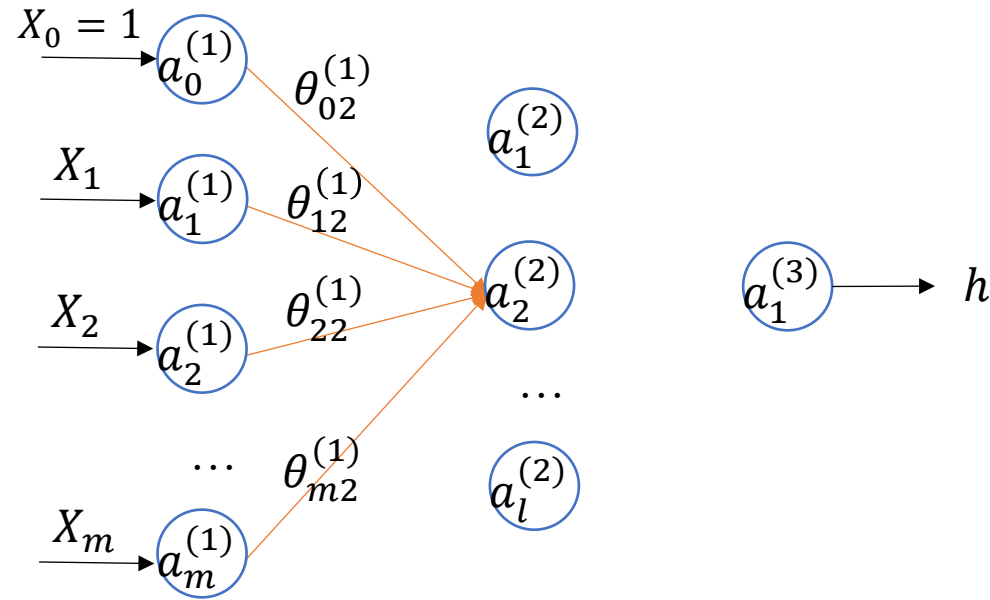
$$a_1^{(2)} = g(\theta_{p1}^{(1)} a_p^{(1)}), \quad p = 0..m$$

$$Z_1^{(2)} = \theta_{p1}^{(1)} a_p^{(1)}$$

* Используется правило Эйнштейна (сумм. по одинак.)

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: $Y=\{0;1\}$



$\theta_{pq}^{(1)}$ — весовые коэффициенты 1 слоя

p — номер нейрона предыдущего слоя;

q — номер нейрона текущего слоя;

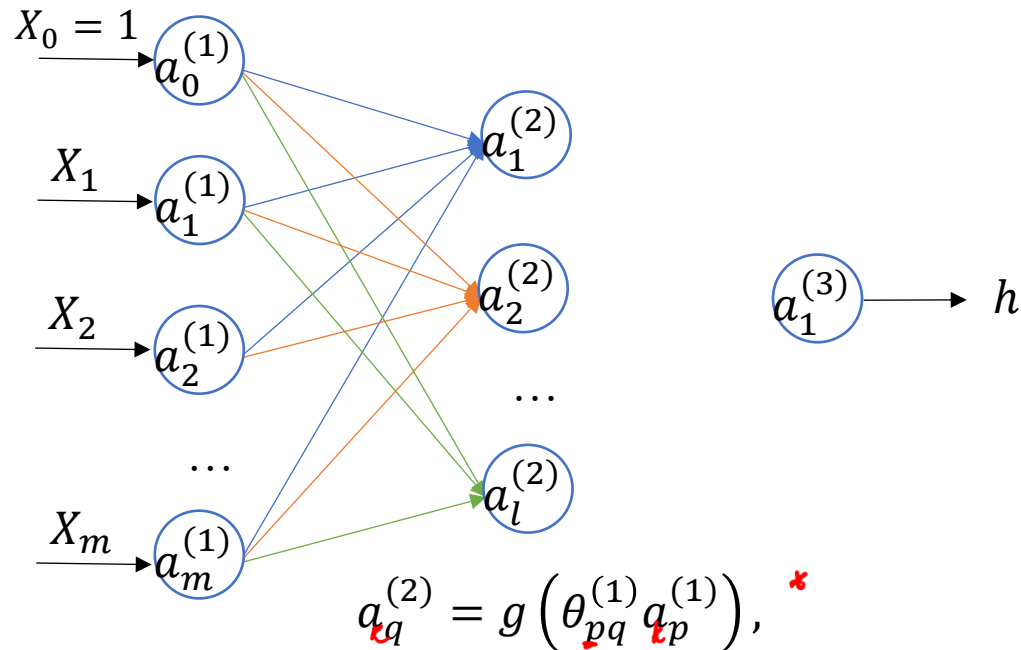
$$a_2^{(2)} = g(\theta_{02}^{(1)} a_0^{(1)} + \theta_{12}^{(1)} a_1^{(1)} + \dots + \theta_{m2}^{(1)} a_m^{(1)})$$

$$a_2^{(2)} = g(\theta_{p2}^{(1)} a_p^{(1)}), \quad p = 0..m$$

$$Z_2^{(2)} = \theta_{p2}^{(1)} a_p^{(1)}$$

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: $Y=\{0;1\}$.



* Иссл. е. Лурье

$p=0..m$ — номер нейрона предыдущего слоя;
 $q=1..l$ — номер нейрона текущего слоя;
 l — кол-во нейронов текущего слоя;

$\theta_{[(m+1) \times l]}^{(1)} = (\theta_{pq}^{(1)})$ — весовые коэффициенты от 1 слоя ко 2му

$$\theta^{(1)} = \begin{pmatrix} \theta_{01}^{(1)} & \theta_{02}^{(1)} & \dots & \theta_{0l}^{(1)} \\ \theta_{11}^{(1)} & \theta_{12}^{(1)} & \dots & \theta_{1l}^{(1)} \\ \dots & \dots & \dots & \dots \\ \theta_{m1}^{(1)} & \theta_{m2}^{(1)} & \dots & \theta_{ml}^{(1)} \end{pmatrix}$$

a_{pq}, θ_{pq}

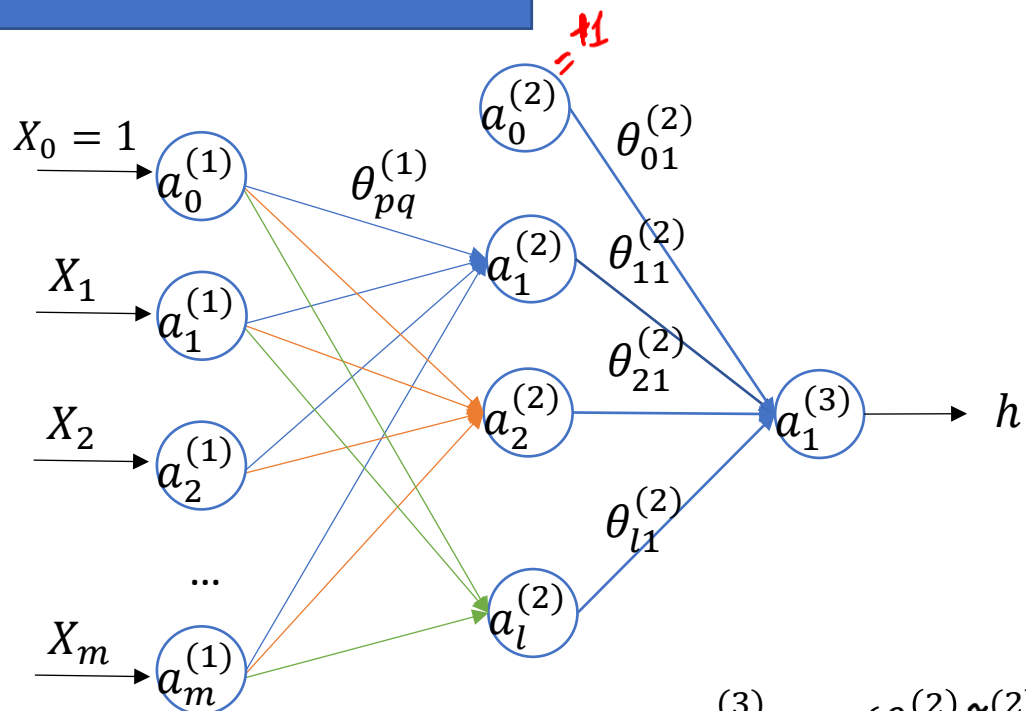
$$a^{(2)} = g(a^{(1)} \theta^{(1)}),$$

$[n \times l] \quad [n \times (m+1)] \quad [(m+1) \times l]$

$g(Z^{(2)})$ — например, логистическая функция активации

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: $Y=\{0;1\}$.



$\theta_{[(l+1) \times 1]}^{(2)}$ – весовые коэффициенты 2 слоя:

$$\theta^{(2)} = \begin{pmatrix} \theta_{01}^{(2)} \\ \theta_{11}^{(2)} \\ \dots \\ \theta_{l1}^{(2)} \end{pmatrix}$$

$\tilde{a}_{[n \times (l+1)]}^{(2)}$ – нейроны второго слоя, включая “+1”:

$$\tilde{a}_{[n \times (l+1)]}^{(2)} = \begin{pmatrix} 1 & a_{11}^{(2)} & \dots & a_{1l}^{(2)} \\ 1 & a_{21}^{(2)} & & a_{2l}^{(2)} \\ \dots & & & \\ 1 & a_{n1}^{(2)} & \dots & a_{nl}^{(2)} \end{pmatrix}$$

$$a_1^{(3)} = g(\theta_{01}^{(2)} \tilde{a}_{10}^{(2)} + \theta_{11}^{(2)} \tilde{a}_{11}^{(2)} + \dots + \theta_{l1}^{(2)} \tilde{a}_{1l}^{(2)})$$

$$a_1^{(3)} = g(\theta_{p1}^{(2)} \tilde{a}_{1p}^{(2)}), \quad p = 0..l$$

$$a^{(3)} = g(\tilde{a}^{(2)} \theta^{(2)}),$$

$[n \times 1]$ $[n \times (l+1)]$ $[(l+1) \times 1]$

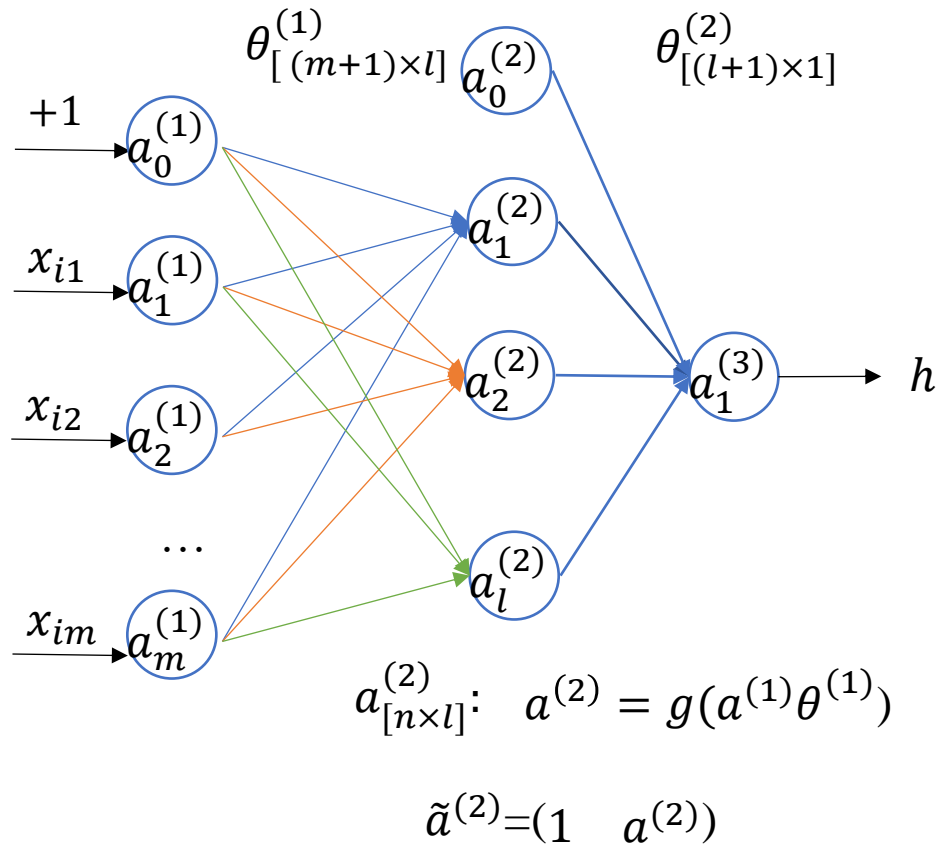
p – номер нейрона предыдущего слоя;

q – номер нейрона текущего слоя;

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: $Y=\{0;1\}$.

Процедура прямого распространения (Forward propagation). Обобщение формул



$$a_{[n \times (m+1)]}^{(1)} = \chi$$

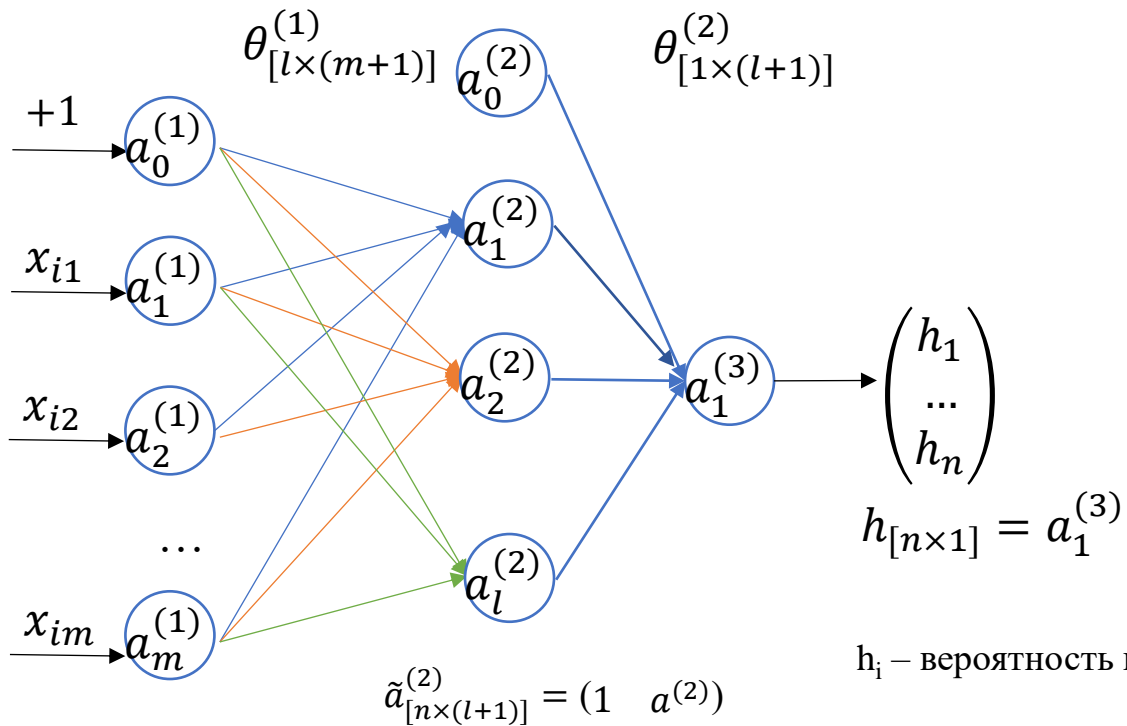
$$\theta^{(1)} = \begin{pmatrix} \theta_{01}^{(1)} & \theta_{02}^{(1)} & \dots & \theta_{0l}^{(1)} \\ \theta_{11}^{(1)} & \theta_{12}^{(1)} & \dots & \theta_{1l}^{(1)} \\ \dots & \dots & \dots & \dots \\ \theta_{m1}^{(1)} & \theta_{m2}^{(1)} & \dots & \theta_{ml}^{(1)} \end{pmatrix}$$

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: $Y=\{0;1\}$.

Процедура прямого распространения (Forward propagation). Обобщение формул



$$a_{[n \times 1]}^{(3)}: a^{(3)} = g(\tilde{a}^{(2)} \theta^{(2)})$$

$$\tilde{a}_{[n \times (l+1)]}^{(2)} = \begin{pmatrix} 1 & a_{11}^{(2)} & \dots & a_{1l}^{(2)} \\ 1 & a_{21}^{(2)} & & a_{2l}^{(2)} \\ \dots & & & \\ 1 & a_{n1}^{(2)} & \dots & a_{nl}^{(2)} \end{pmatrix} \quad \theta^{(2)} = \begin{pmatrix} \theta_{01}^{(2)} \\ \theta_{11}^{(2)} \\ \dots \\ \theta_{l1}^{(2)} \end{pmatrix}$$

h_i – вероятность принадлежности i -го объекта классу «1»

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: $Y=\{0;1\}$.

Поиск весовых коэффициентов. Функция ошибки. Процедура прямого распространения ошибки.

Cross Entropy loss

$$J(\theta^{(1)}, \theta^{(2)}) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(\underline{h_i})) + (1 - y_i) \ln(1 - \underline{h_i})) \rightarrow \min$$

Матричный вид:

$$J(\theta^{(1)}, \theta^{(2)}) = -\frac{1}{n} (Y^T \ln(h) + (I - Y)^T \ln(I - h)) \rightarrow \min,$$

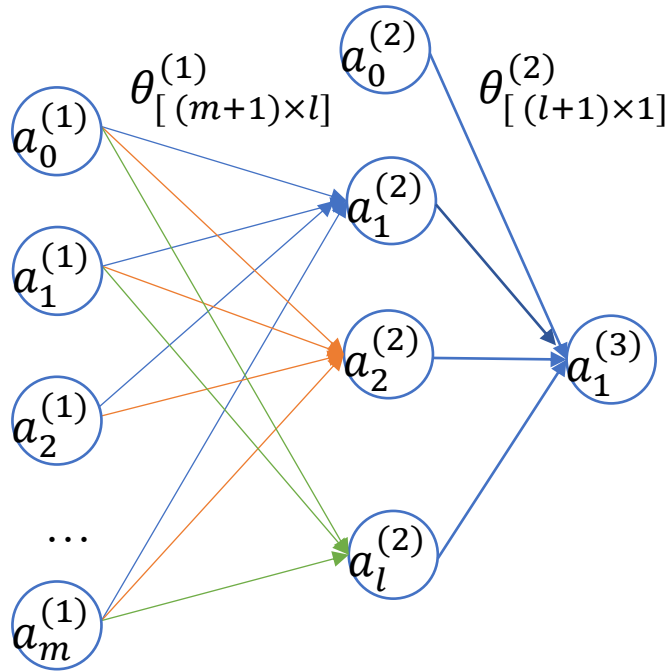
где $I = (1)_{[n \times 1]}$

С учетом регуляризации:

$$J(\theta^{(1)}, \theta^{(2)}) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_i)) + (1 - y_i) \ln(1 - h_i)) + \frac{\lambda}{2n} \left(\sum_{p=1}^m \sum_{q=1}^l \theta_{pq}^{(1)} \theta_{pq}^{(1)} + \sum_{p=1}^l \theta_{p1}^{(2)} \theta_{p1}^{(2)} \right) \rightarrow \min$$

Матричный вид:

$$J(\theta^{(1)}, \theta^{(2)}) = -\frac{1}{n} (Y^T \ln(h) + (I - Y)^T \ln(I - h)) + \frac{\lambda}{2n} ((\theta^{(1)})^T \theta^{(1)} + (\theta^{(2)})^T \theta^{(2)}) \rightarrow \min$$

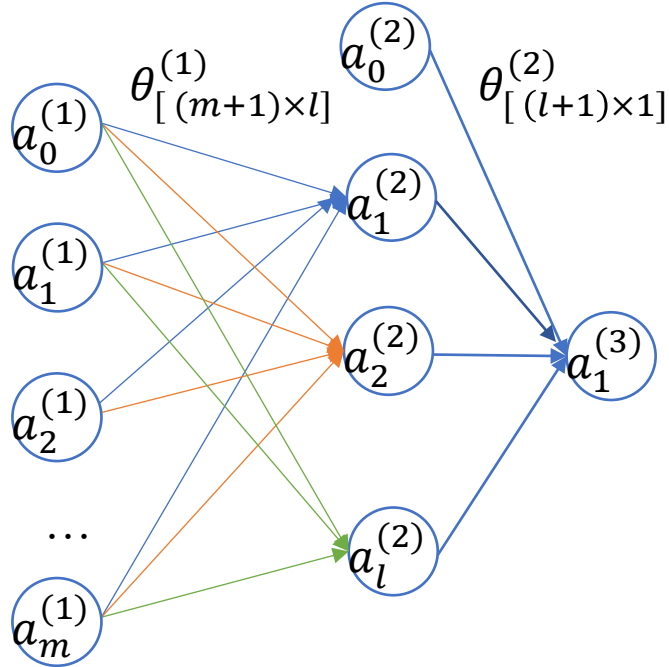


Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: $Y = \{0; 1\}$.

Поиск весовых коэффициентов. Функция ошибки. Процедура прямого распространения ошибки.



$$a_{[n \times (m+1)]}^{(1)} = \chi$$

Процедура прямого распространения ошибки:

- выбор начальных значений вес. коэф-в (нач. точка):

$$\theta^{(1)} = \theta_{\text{нач}}^{(1)}, \theta^{(2)} = \theta_{\text{нач}}^{(2)} \text{ (не нулевые)}$$

- вычисление значений в каждом нейроне на скрытом слое:

$$Z^{(2)} = a^{(1)} \theta^{(1)}, a^{(2)} = g(Z^{(2)})$$

- вычисление значений на выходном слое:

$$\tilde{a}^{(2)} = (1 \quad a^{(2)}), Z^{(3)} = \tilde{a}^{(2)} \theta^{(2)},$$

$$a^{(3)} = g(Z^{(3)}), h = a^{(3)}$$

- вычисление функции ошибки:

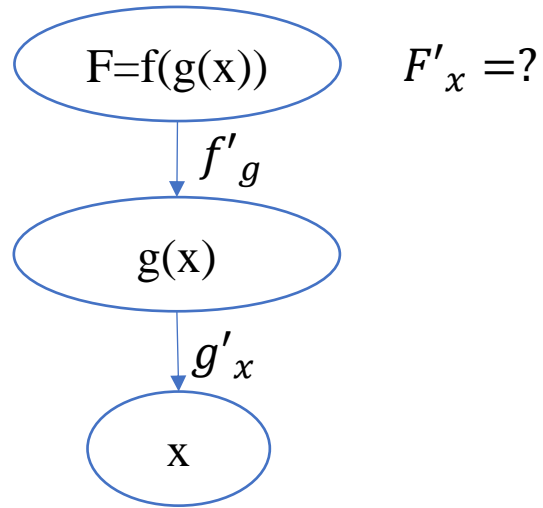
$$J = -\frac{1}{n} (Y^T \ln(h) + (I - Y)^T \ln(I - h))$$

Введение в ИНС

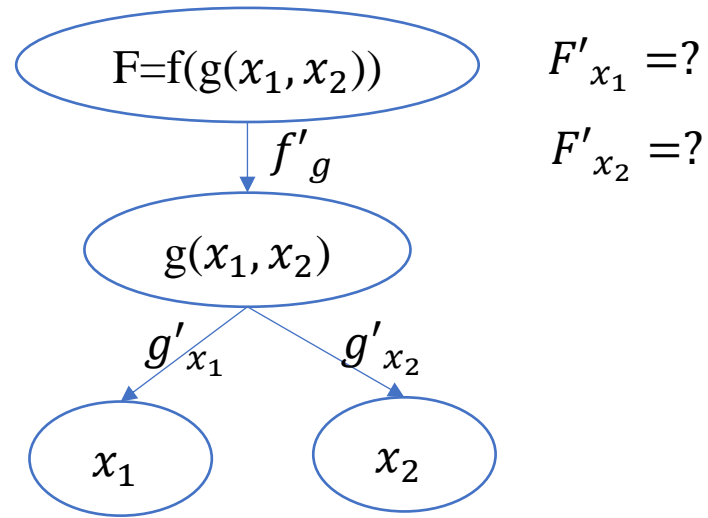
Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

Правило дифференцирования сложной функции:



$$F'_x = f'_g \times g'_x$$



$$F'_{x_1} = f'_g \times g'_{x_1} \quad F'_{x_2} = f'_g \times g'_{x_2}$$

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

Логистическая функция активации - сигмоида:

$$g(z) = (1 + e^{-z})^{-1}$$

Дифференцирование сигмоиды:

$$g(z) = (1 + e^{-z})^{-1}$$

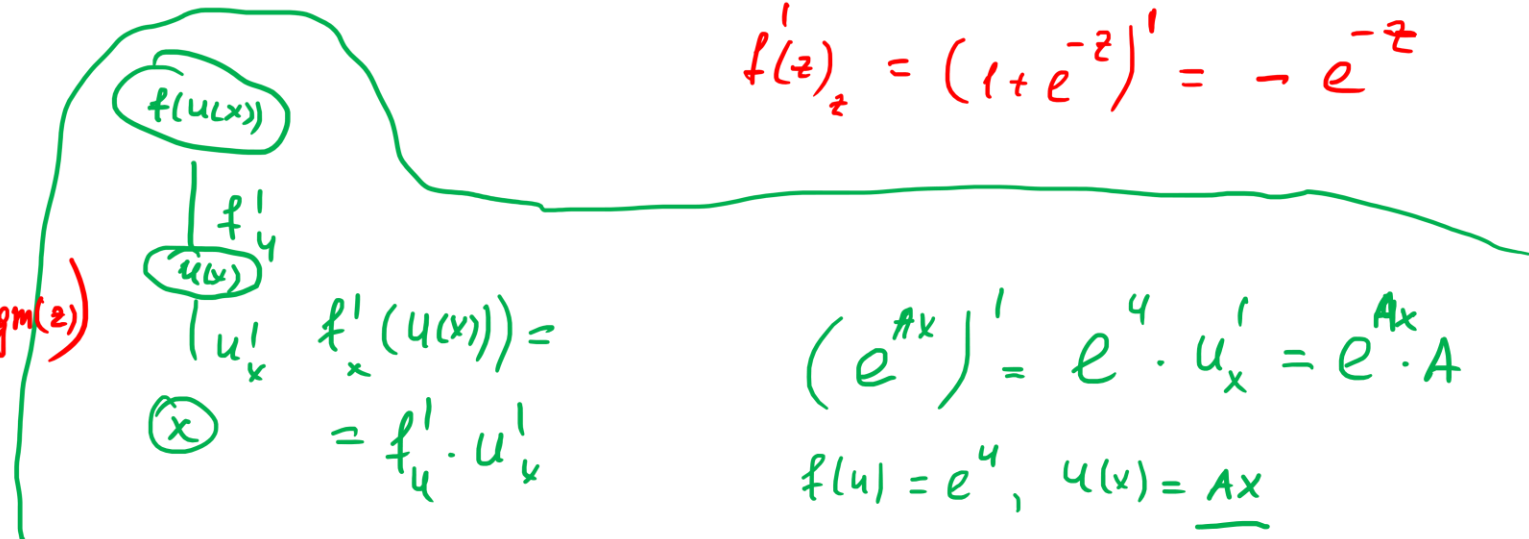
$$\begin{aligned} g'(z) &= g'_f \cdot f'_z = -\frac{1}{f^2} \cdot (-e^{-z}) = \\ &= \frac{(e^{-z} + 1) - 1}{(1 + e^{-z})^2} = \frac{1}{1 + e^{-z}} - \frac{1}{(1 + e^{-z})^2} = \\ g'(z) &= (1 - g(z))g(z) \\ &= \frac{1}{1 + e^{-z}} \left(1 - \frac{1}{1 + e^{-z}} \right) = \text{sigm}(z)(1 - \text{sigm}(z)) \end{aligned}$$

$$g(f(z))'_z = g'_f \cdot f'_z$$

$$\frac{f(z)}{g(f)} = \frac{1 + e^{-z}}{f^{-1}}$$

$$g'_f = -1 \cdot f^{-2} = -\frac{1}{f^2}$$

$$f'_z = (1 + e^{-z})' = -e^{-z}$$

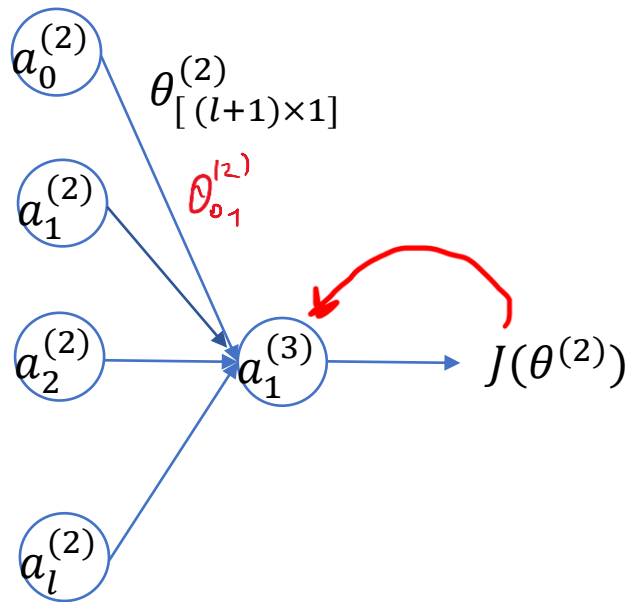


Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

$$J(\theta^{(1)}, \theta^{(2)}) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(a^{(3)}) + (1 - y_i) \ln(1 - a^{(3)})) \rightarrow \min$$



$$J(a^{(3)}) = -\frac{1}{n} (y \ln(a^{(3)}) + (1 - y) \ln(1 - a^{(3)}))$$

$$(h_2(x))' = \frac{1}{x}$$

$$J'_{a^{(3)}} = -\frac{1}{n} \left(\frac{y}{a^{(3)}} + \frac{-(1-y)}{1-a^{(3)}} \right) = \frac{1}{n} \cdot \frac{a^{(3)} - y}{a^{(3)}(1-a^{(3)})}$$

$$J'_{a^{(3)}} = \frac{1}{n} \frac{a^{(3)} - y}{a^{(3)}(1-a^{(3)})}$$

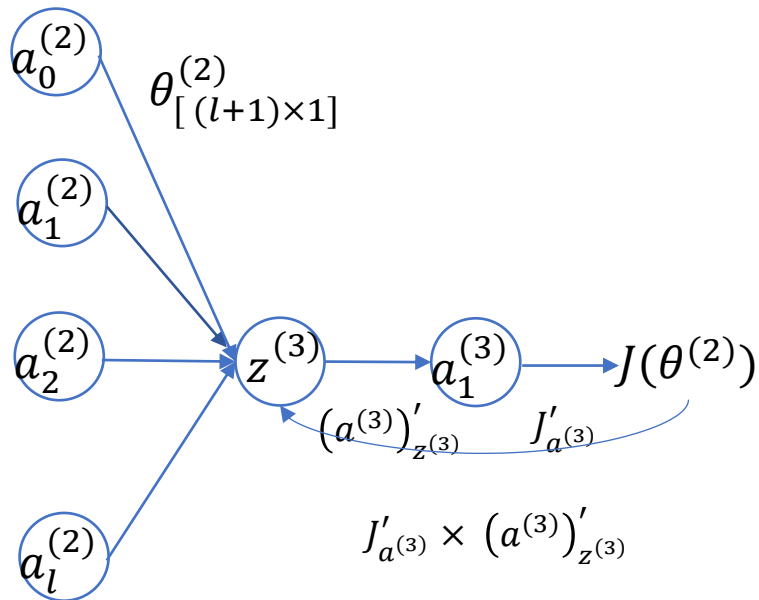
Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

Дифференцирование $J(z)$:

$$J(z) = J(a^{(3)}(z^{(3)}))$$

$$(a^{(3)}(z^{(3)}))' = (1 - a^{(3)}(z)) a^{(3)}(z) \quad J'_{a^{(3)}} = \frac{1}{n} \frac{a^{(3)} - y}{a^{(3)}(1 - a^{(3)})}$$



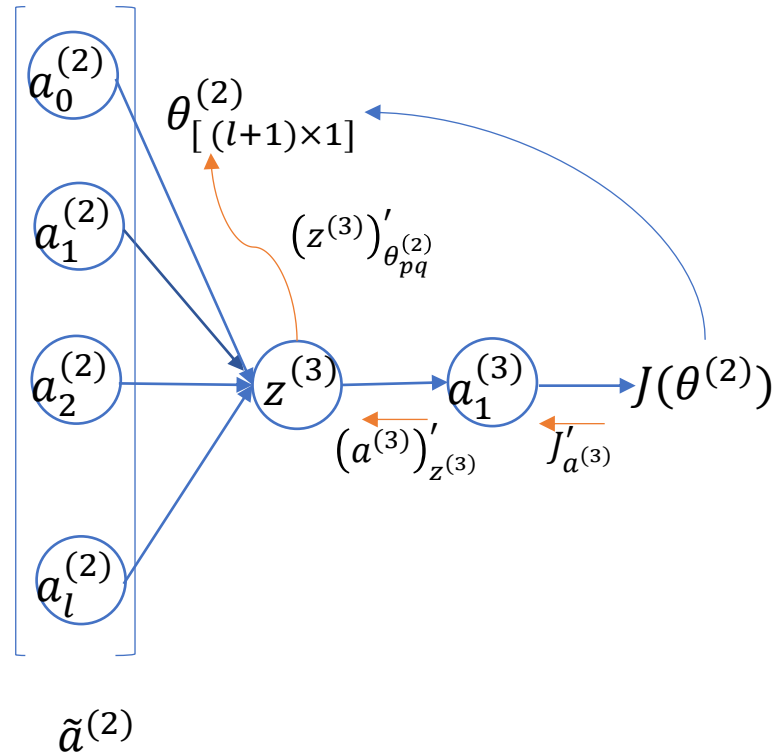
$$J'_{z^{(3)}} = J'_{a^{(3)}} \times (a^{(3)})'_{z^{(3)}} = \frac{1}{n} (a^{(3)} - y) = \delta^{(3)}$$

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

$$J'_{z^{(3)}} = \delta^{(3)}$$



$$J(a^{(3)}(z^{(3)}(\theta^{(2)}))) = -\frac{1}{n}(y \ln(a^{(3)}) + (1-y)(1 - \ln(a^{(3)})))$$

$$z^{(3)} = \tilde{a}^{(2)} \theta^{(2)}$$

$$\frac{\partial J}{\partial \theta_{p1}^{(2)}} = \frac{\partial J}{\partial z^{(3)}} \frac{\partial z_j^{(3)}}{\partial \theta_{p1}^{(2)}} = \delta^{(3)} \tilde{a}_p^{(2)}, \quad p = 0, \dots, l$$

$$\frac{\partial J}{\partial \theta_{p1}^{(2)}} = \delta_j^{(3)} \tilde{a}_{jp}^{(2)}, \quad j = 1, \dots, n, p = 0, \dots, l,$$

$$\text{где } \delta_{[n \times 1]}^{(3)} = \frac{a^{(3)} - y}{n}$$

(1)

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

$$J(\theta^{(1)}, \theta^{(2)}) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(a^3)) + (1 - y_i) \ln(1 - a^3)) \rightarrow \min$$

$$a'(z) = a(z)(1 - a(z))$$

$$J(\theta^{(1)}) = J(a^{(2)}(z^{(2)}(\theta^{(1)}))), \quad \underline{z^{(2)} = a^{(1)}\theta^{(1)}}$$

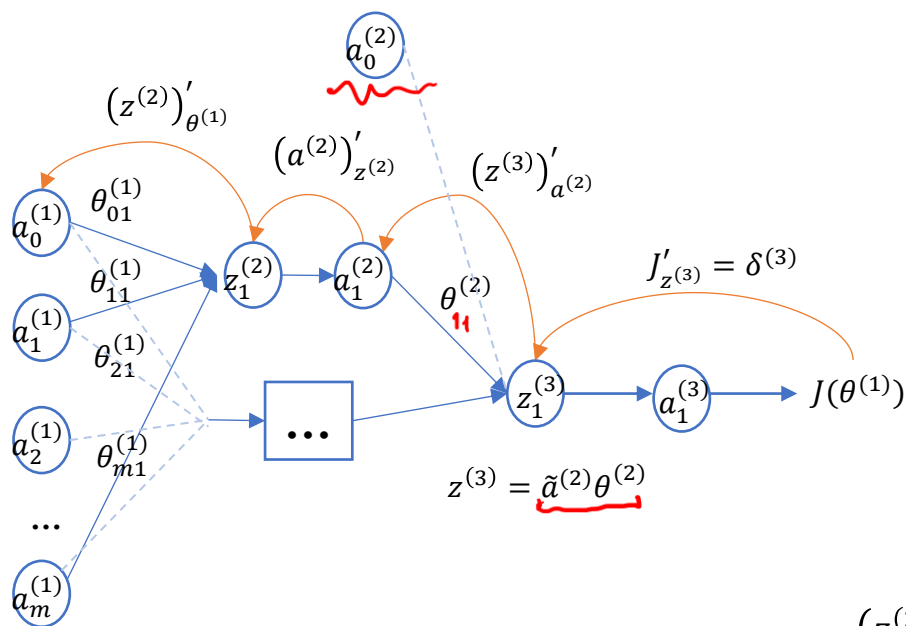


Рис.1. Фрагмент ИНС

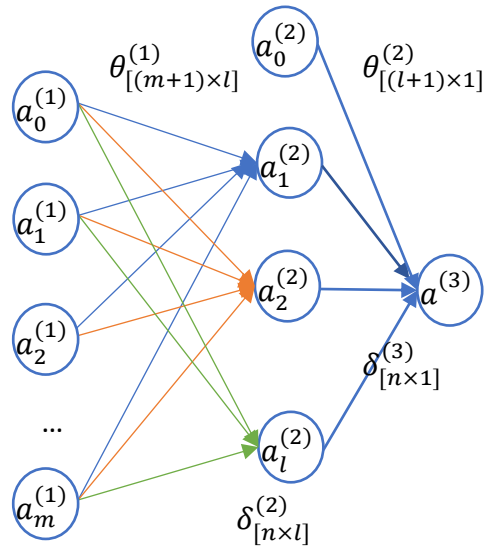
$$\left. \begin{aligned} J'_{z^{(3)}} &= \delta^{(3)} \\ (z^{(3)})'_{a^{(2)}} &= \theta^{(2)} \end{aligned} \right\} J'_{a^{(2)}} = \delta^{(3)} \theta^{(2)} \quad \text{#без } a_0^{(2)}$$

$$\left. \begin{aligned} (a^{(2)})'_{z^{(2)}} &= a^{(2)}(1 - a^{(2)}) \\ (z^{(2)})'_{\theta^{(1)}} &= a^{(1)} \end{aligned} \right\} J'_{z^{(2)}} = \delta^{(3)} \theta^{(2)} a^{(2)} (1 - a^{(2)}) \quad \text{#без } \delta_0^{(2)}$$

$$J'_{z^{(2)}} = \underbrace{\delta^{(3)} \theta^{(2)}}_{\hat{\delta}^{(2)}} a^{(2)} (1 - a^{(2)}) \underbrace{\tilde{a}^{(1)}}_{\delta_{[n \times l]}^{(2)}}$$

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.



$$J(\theta^{(1)}, \theta^{(2)}) = -\frac{1}{n} (Y^T \ln(a^{(3)}) + (I - Y)^T \ln(I - a^{(3)}))$$

$$\frac{\partial J}{\partial \theta_{pq}^{(1)}} = \delta_{iq}^{(2)} \tilde{a}_{ip}^{(1)}, \quad i=1..n, \quad q=1..l, \quad p=0..m,$$

$$\text{где } \delta_{iq}^{(2)} = \hat{\delta}_{iq}^{(2)} a_{iq}^{(2)} (1 - a_{iq}^{(2)}), \quad q=1..l, \quad i=1..n, \text{ #без } \hat{\delta}_0^{(2)}$$

$$\hat{\delta}_{iq}^{(2)} = \delta_i^{(3)} \theta_{q1}^{(2)}, \quad i=1..n, \quad q=0..l$$

(2)

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

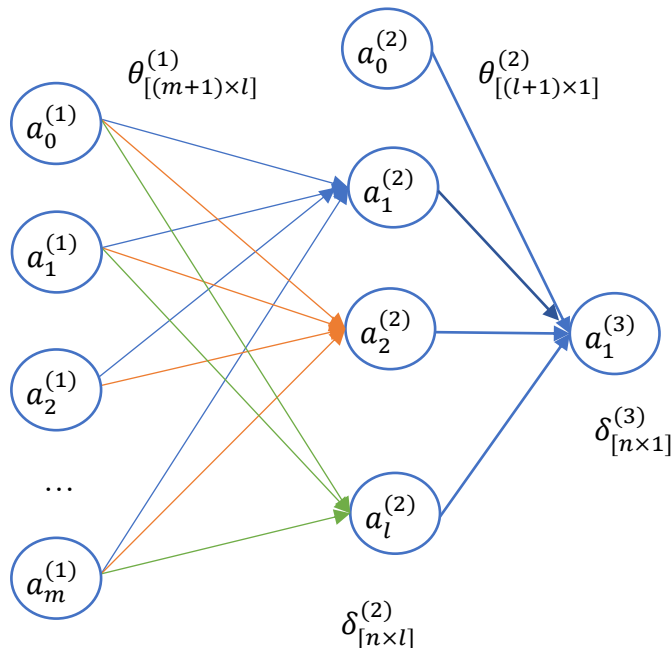
Поиск весовых коэффициентов. Метод обратного распространения ошибки.

Ошибки на каждом слое:

$$\delta_i^{(3)} = \frac{a_i^{(3)} - y_i}{n}, \quad i = 1..n \quad (3)$$

$$\hat{\delta}_{iq}^{(2)} = \delta_i^{(3)} \theta_{q1}^{(2)}, \quad i = 1..n, \quad q = 0..l$$

$$\delta_{iq}^{(2)} = \hat{\delta}_{iq}^{(2)} a_{iq}^{(2)} (1 - a_{iq}^{(2)}), \quad q = 1..l, \quad i=1..n \text{ \textcolor{red}{#без } } \hat{\delta}_0^{(2)}$$



$$a(z) = (1 + e^{-z})^{-1}$$

Компоненты градиента функции ошибки:

$$\frac{\partial J}{\partial \theta_{p1}^{(2)}} = \delta_i^{(3)} \tilde{a}_{ip}^{(2)}, \quad p = 0, \dots, l, \quad i = 1, \dots, n$$

$$\frac{\partial J}{\partial \theta_{pq}^{(1)}} = \delta_{iq}^{(2)} \tilde{a}_{ip}^{(1)}, \quad p = 0..m, \quad q = 1..l, \quad i = 1..n$$

где $\delta_{iq}^{(2)} = \hat{\delta}_{iq}^{(2)} a_{iq}^{(2)} (1 - a_{iq}^{(2)}), \quad q = 1..l, \quad i=1..n \text{ \textcolor{red}{#без } } \hat{\delta}_0^{(2)}$

(4)

Архитектура сетей прямого распространения. Задача многоклассовой классификации.

Компоненты градиента функции ошибки:

$$\frac{\partial J}{\partial \theta_{pq}^{(2)}} = \delta_{iq}^{(3)} \tilde{a}_{ip}^{(2)}, p = 0, \dots, l, q = 1, \dots, k, i = 1, \dots, n \quad (5)$$

$$\frac{\partial J}{\partial \theta_{pq}^{(1)}} = \delta_{iq}^{(2)} \tilde{a}_{ip}^{(1)}, p = 0, \dots, m, q = 1, \dots, l, i = 1, \dots, n$$

где $\delta_{iq}^{(2)} = \hat{\delta}_{iq}^{(2)} a_{iq}^{(2)} (1 - a_{iq}^{(2)})$, $q = 1, \dots, l, i = 1, \dots, n$ **#без $\hat{\delta}_0^{(2)}$**

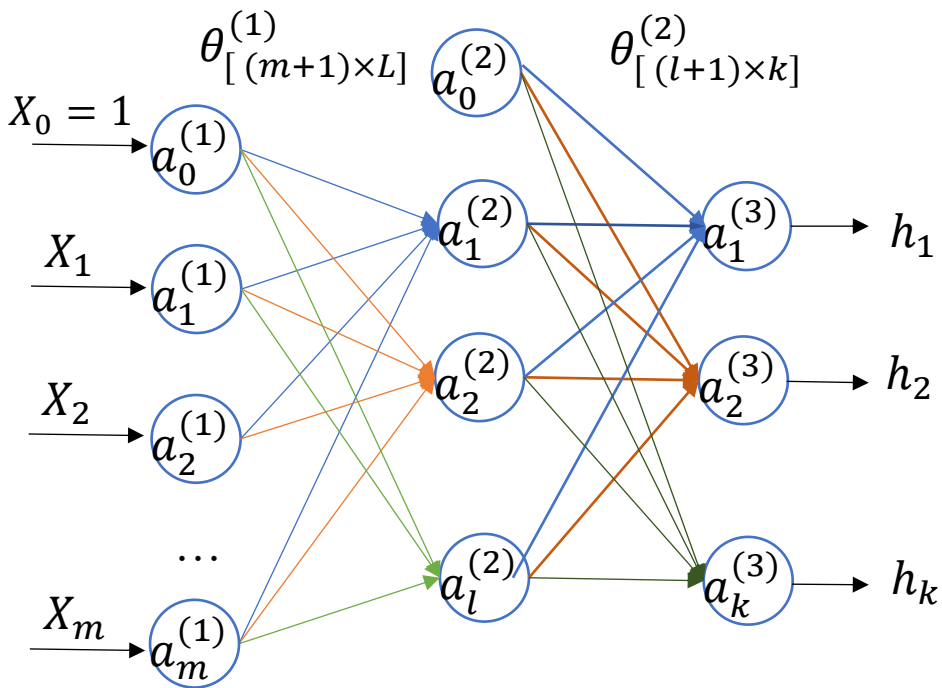
Компоненты градиента функции ошибки в матричном виде:

$$\frac{\partial J}{\partial \theta^{(2)}} = (\tilde{a}^{(2)})^T \delta^{(3)},$$

$$\frac{\partial J}{\partial \theta^{(1)}} = (\tilde{a}^{(1)})^T \delta^{(2)}, \quad (5^*)$$

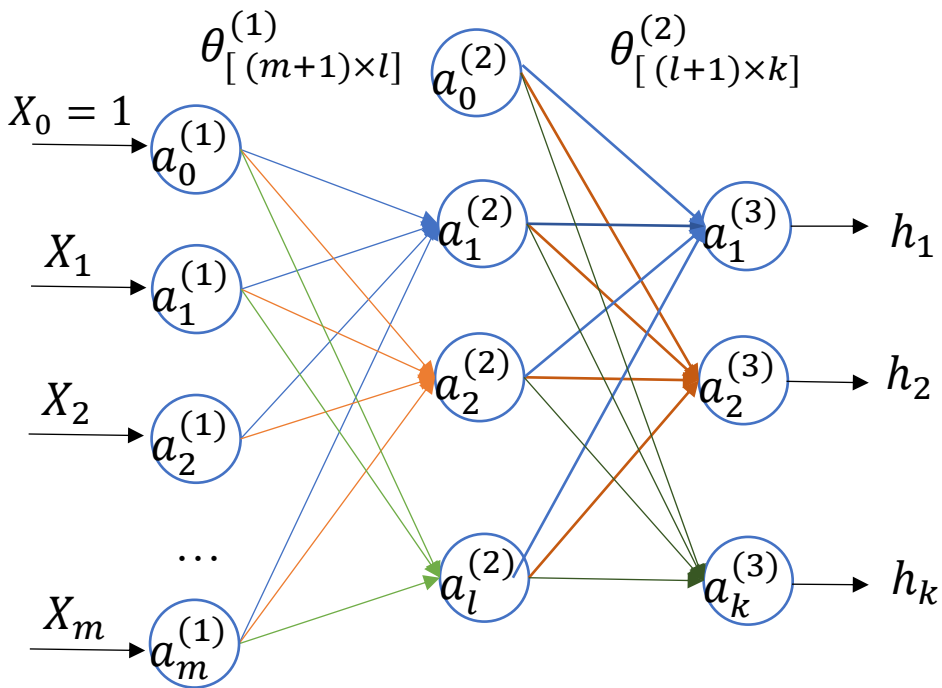
где $\delta^{(2)} = \hat{\delta}^{(2)} \cdot a^{(2)} \cdot (1 - a^{(2)})$, **#без $\hat{\delta}_0^{(2)}$**

\cdot – означает покомпонентное произведение



$$a(z) = (1 + e^{-z})^{-1}$$

Архитектура сетей прямого распространения. Задача многоклассовой классификации.



$$a(z) = (1 + e^{-z})^{-1}$$

Обозначение	Описание
$\theta_{[(m+1) \times l]}^{(1)} = (\theta_{pq}^{(1)})$	весовые коэффициенты от 1 слоя ко 2му
$\theta_{[(l+1) \times k]}^{(2)} = (\theta_{pq}^{(2)})$	весовые коэффициенты от 2 слоя к 3му
$a_{[n \times l]}^{(2)} = (a_{ip}^{(2)})$	матрица со значениями в нейронах скрытого слоя для n объектов
$\tilde{a}_{[n \times (l+1)]}^{(2)} = (1 \quad a^{(2)})$	матрица значений в нейронах скрытого слоя, включая bias
$a_{[n \times k]}^{(3)} = (a_{ip}^{(3)})$	матрица со значениями в нейронах выходного слоя для n объектов
m	кол-во признаков;
l	кол-во нейронов скрытого слоя, не считая bias
k	кол-во нейронов выходного слоя (кол-во классов)
$a(z) = (1 + e^{-z})^{-1}$	функция активации на скрытом и выходном слое

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

Компоненты градиента:

$$\frac{\partial J}{\partial \theta_{01}^{(2)}} = \delta_j^{(3)} \tilde{a}_{jq}^{(2)}, j = 1, \dots, n,$$

$$\frac{\partial J}{\partial \theta_{q1}^{(2)}} = \left(\delta_j^{(3)} \tilde{a}_{jq}^{(2)} + \frac{\lambda}{n} \theta_{q1}^{(2)} \right), q = 1, \dots, l, j = 1, \dots, n,$$

$$\frac{\partial J}{\partial \theta_{0q}^{(1)}} = \delta_{pi}^{(2)} \tilde{a}_{iq}^{(1)}, i = 0..m, \text{ // без } \delta_0^{(2)}$$

$$\frac{\partial J}{\partial \theta_{pq}^{(1)}} = \left(\delta_{pi}^{(2)} \tilde{a}_{iq}^{(1)} + \frac{\lambda}{n} \theta_{pq}^{(1)} \right), i = 0..m, p = 1..l, \text{ // без } \delta_0^{(2)}$$

$$\Rightarrow \nabla J = \begin{bmatrix} \frac{\partial J}{\partial \theta_{pq}^{(1)}} \\ \frac{\partial J}{\partial \theta_{pq}^{(2)}} \end{bmatrix}_{[(m+1) \times l + (l+1) \times k]}$$

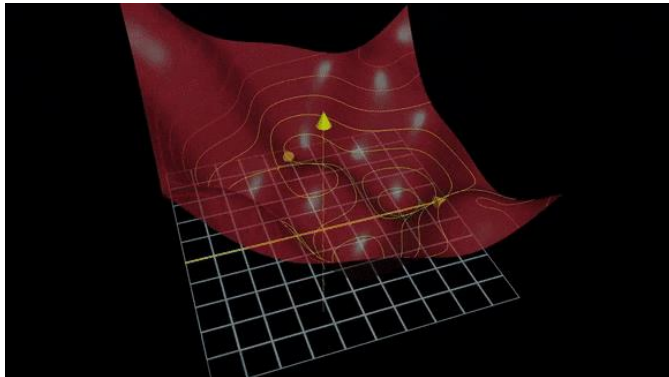
Расчет коэффициентов на каждом шаге $s+1$:

$$\boldsymbol{\theta}^{s+1} = \boldsymbol{\theta}^s - \alpha \nabla J|_s$$

$\boldsymbol{\theta}^0 = (\theta_{pq}^0)$ — начальное значение коэффициентов;

$\nabla J|_s = \left(\frac{\partial J}{\partial \theta_{pq}} \right) |_s$ — компоненты градиента рассчитываются на каждом шаге $s+1$;

α — шаг градиентного спуска (скорость обучения).



Gradient descent intuition

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

С учетом регуляризации:

$$J(\theta^{(1)}, \theta^{(2)}) = -\frac{1}{n} \sum_{i=1}^n (y_i \ln(h_i)) + (1 - y_i) \ln(1 - h_i) + \frac{\lambda}{2n} \left(\sum_{p=1}^m \sum_{q=1}^l \theta_{pq}^{(1)} \theta_{pq}^{(1)} + \sum_{p=1}^l \theta_{p1}^{(2)} \theta_{p1}^{(2)} \right) \rightarrow \min$$

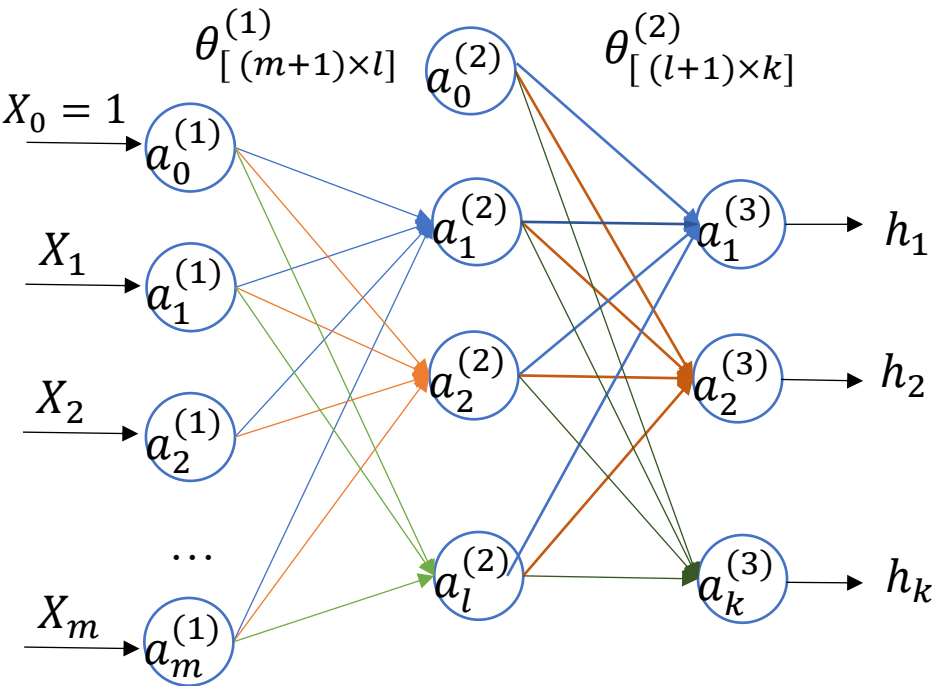
Компоненты градиента

функции ошибки с учетом регуляризации:

$$\begin{aligned} \frac{\partial J}{\partial \theta_{0q}^{(2)}} &= \delta_{iq}^{(3)} \tilde{a}_{i0}^{(2)}, \quad q=1, \dots, k, \quad i=1, \dots, n \\ \frac{\partial J}{\partial \theta_{pq}^{(2)}} &= \delta_{iq}^{(3)} \tilde{a}_{ip}^{(2)} + \frac{\lambda}{n} \theta_{pq}^{(2)}, \quad p=1, \dots, l, \quad q=1, \dots, k, \quad i=1, \dots, n \\ \frac{\partial J}{\partial \theta_{0q}^{(1)}} &= \delta_{iq}^{(2)} \tilde{a}_{i0}^{(1)}, \quad q=1, \dots, l, \quad i=1, \dots, n \\ \frac{\partial J}{\partial \theta_{pq}^{(1)}} &= \delta_{iq}^{(2)} \tilde{a}_{ip}^{(1)} + \frac{\lambda}{n} \theta_{pq}^{(1)}, \quad p=1, \dots, l, \quad q=1, \dots, l, \quad i=1, \dots, n \end{aligned}$$

(6)

где $\delta_{iq}^{(2)} = \hat{\delta}_{iq}^{(2)} a_{iq}^{(2)} (1 - a_{iq}^{(2)})$, $q = 1..l, i=1..n$ *#без $\hat{\delta}_0^{(2)}$*



$$a(z) = (1 + e^{-z})^{-1}$$