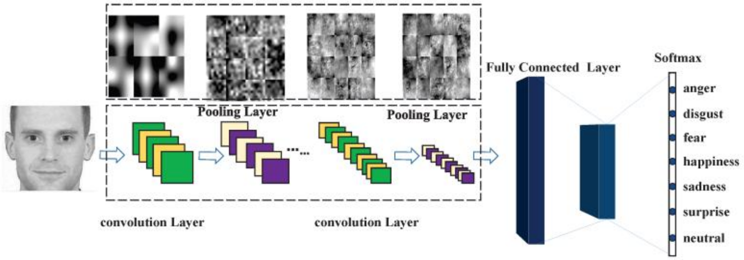
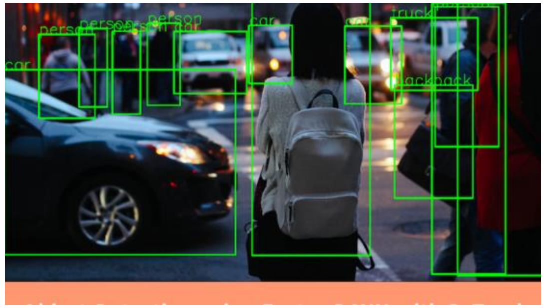
Основы искусственного интеллекта





Лекция 7

Введение в ИНС

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

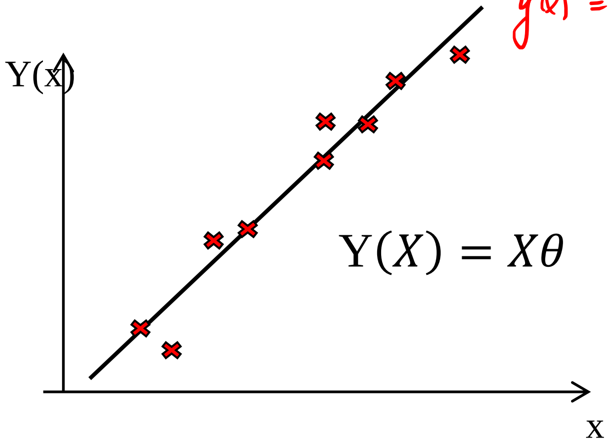
Введение в ИНС

Понятие ИНС

От простого к сложному

Задача аппроксимации

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



$\bar{y}(x) = \theta_0 x_0 + \theta_1 x_1 =$

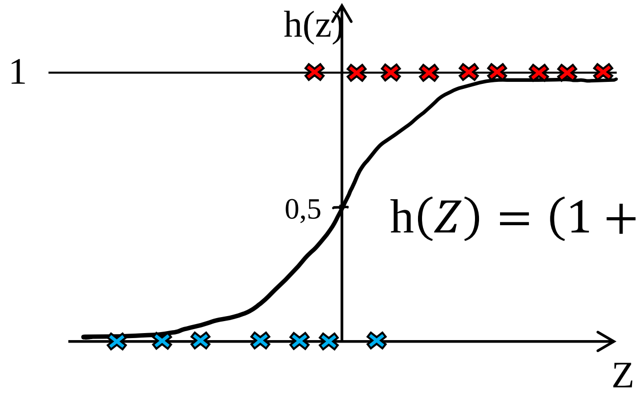
$= \sum \theta_j x_j$

$\mathrm{Y}(X) = X\theta$

Y – непрерывная величина

Задача классификации

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



$\mathrm{h}(Z) = (1 + \exp(-Z))^{-1}$

$Z = X\theta$

Y – дискретная величина

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

$$\theta = \begin{bmatrix} \theta_0 \\ \theta_1 \end{bmatrix} \quad X = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \dots & \dots \\ 1 & x_n \end{bmatrix}$$

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

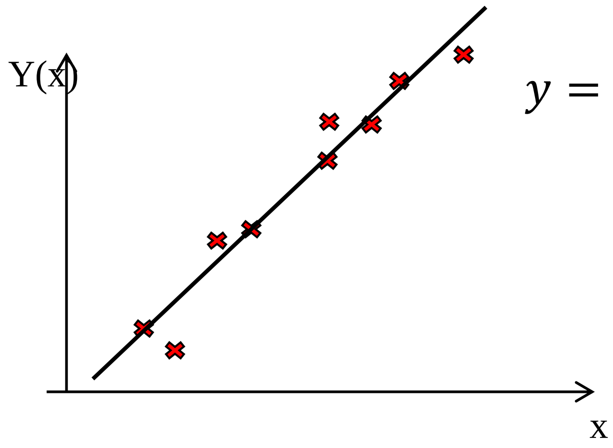
Введение в ИНС

Понятие ИНС

От простого к сложному

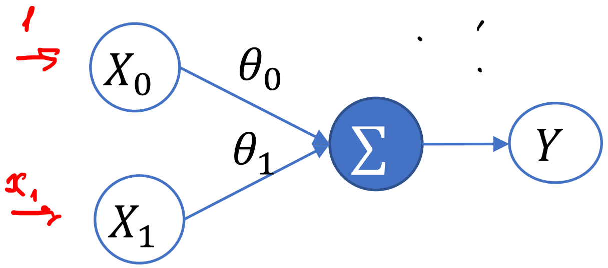
Задача аппроксимации

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



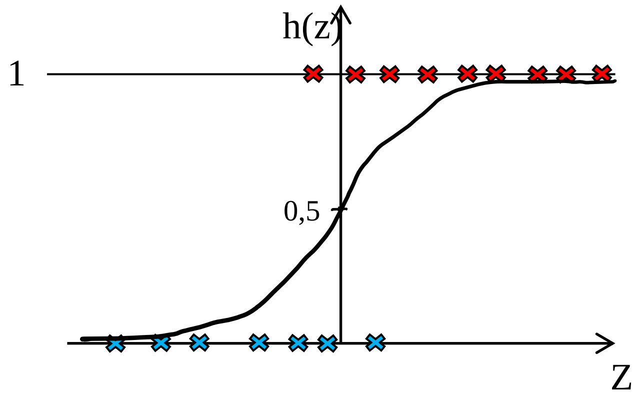
$$y = \begin{bmatrix} y_1 \\ y_2 \\ \\ y_n \end{bmatrix}$$

$$\mathrm{Y}(X) = X\theta$$



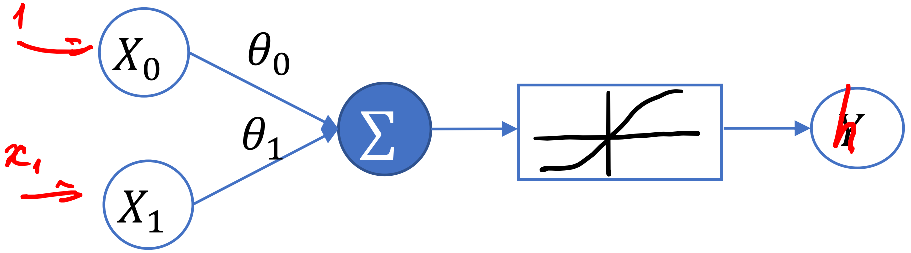
Задача классификации

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



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

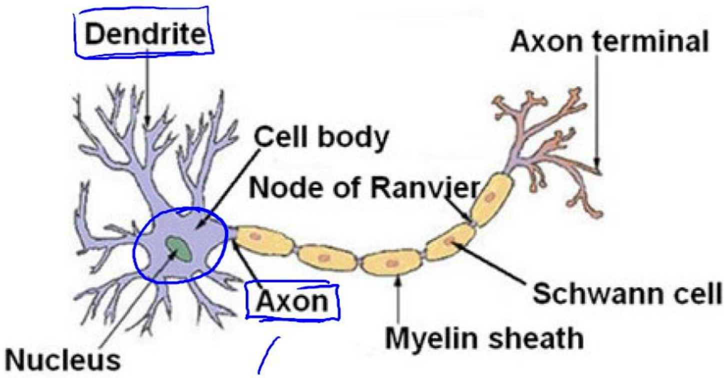
$$\mathrm{Y}(Z) = (1 + \exp(-X\theta))^{-1}$$



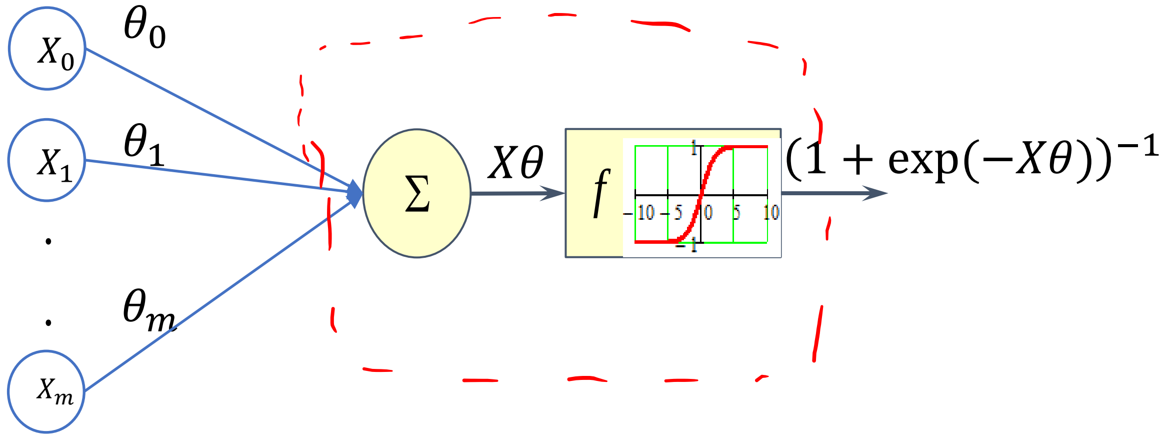
Введение в ИНС

Понятие ИНС. Биологическая аналогия

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



Однослойный перцептрон



дендриты – получают информацию;

тело клетки – обрабатывает информацию;

аксон – передает обработанную информацию другим нейронам;

синапсы – соединяют аксон и дендриты других нейронов.

	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

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

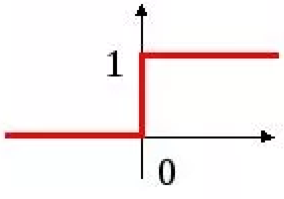
Введение в ИНС

Понятие ИНС

Функции активации

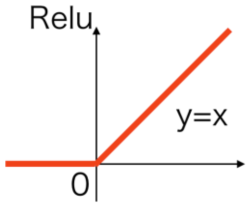
Ступенчатая функция

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



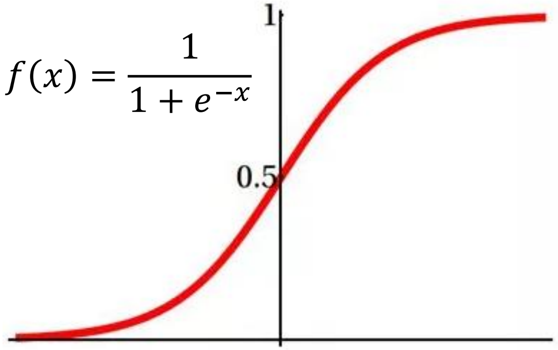
ReLU

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



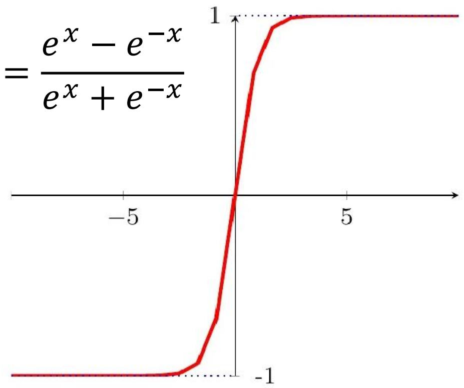
Сигмоида

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



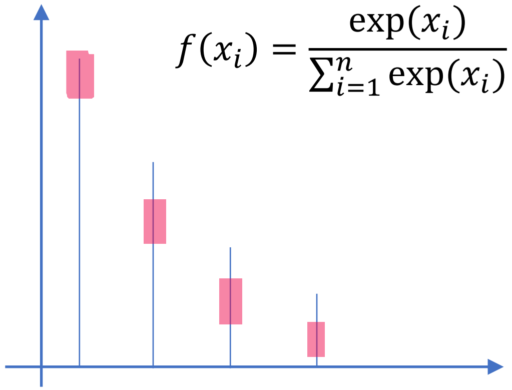
Гиперболический тангенс

$$f(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



softmax

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

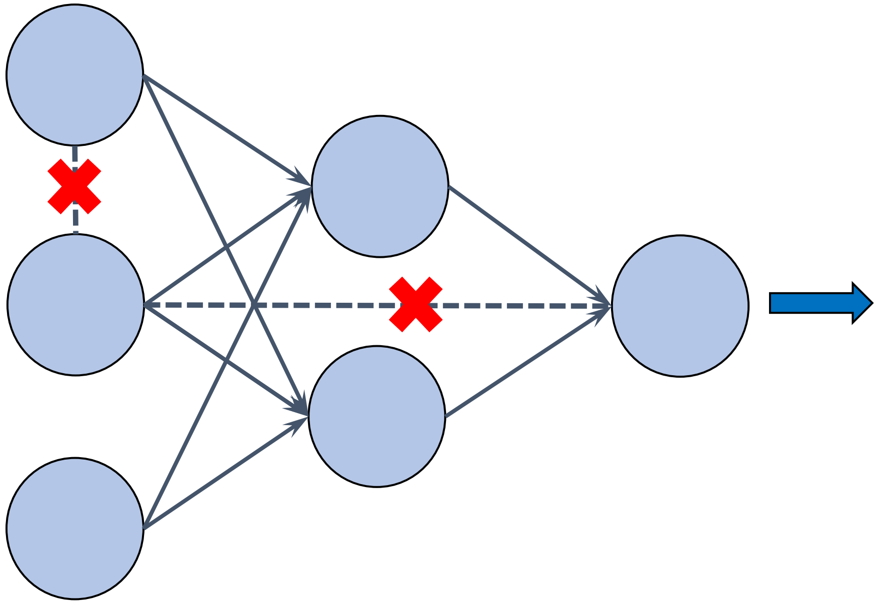


...

Введение в ИНС

Архитектура сетей прямого распространения / Feed forward neural network

Нейроны слоя не связаны.
Нейроны передают информацию только нейронам следующего слоя.
Перепрыгивание через слои запрещено.

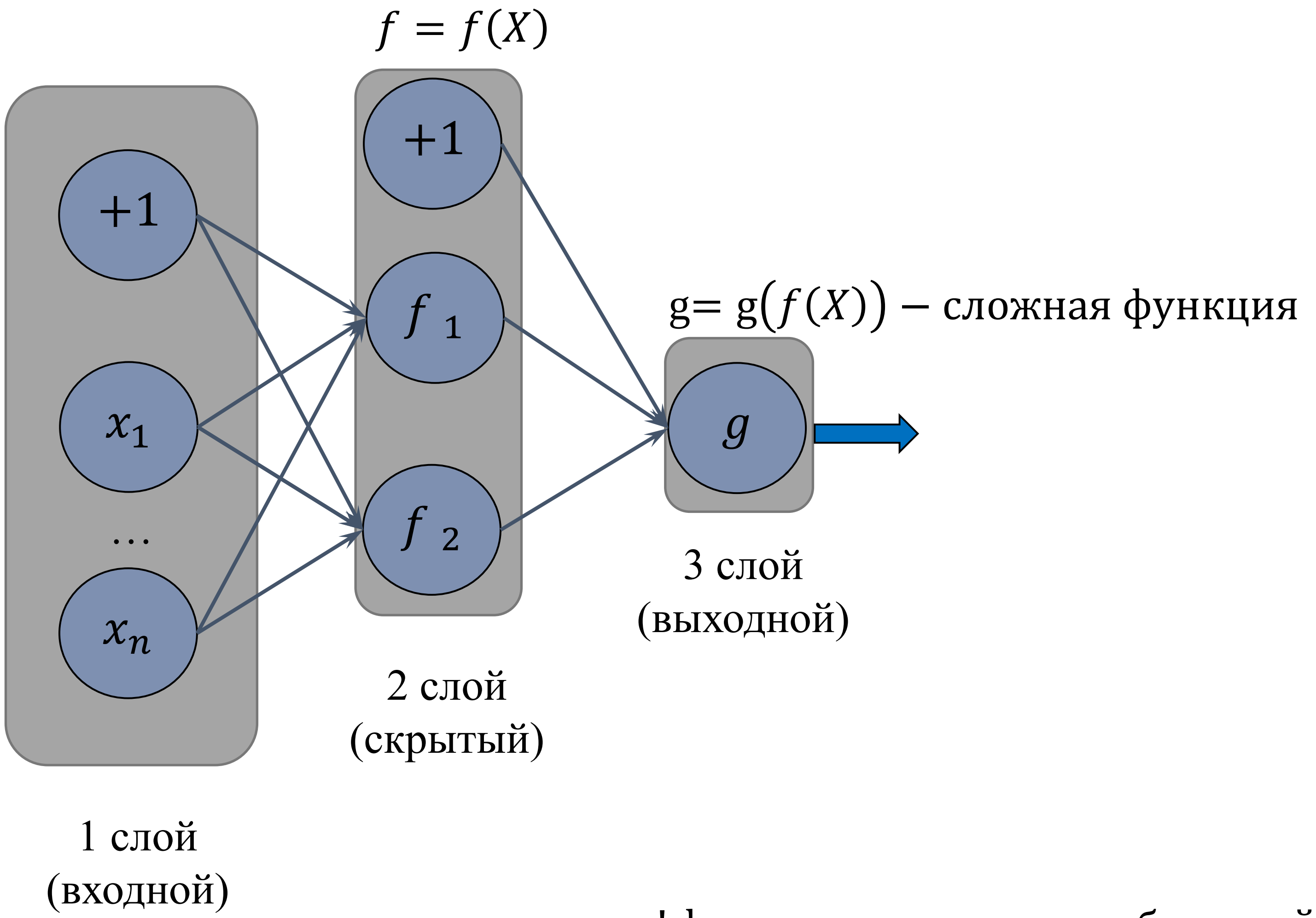


Направление потока информации

Введение в ИНС

Архитектура сетей прямого распространения / Feed forward neural network

Пример.
3х – слойная ИНС прямого распространения

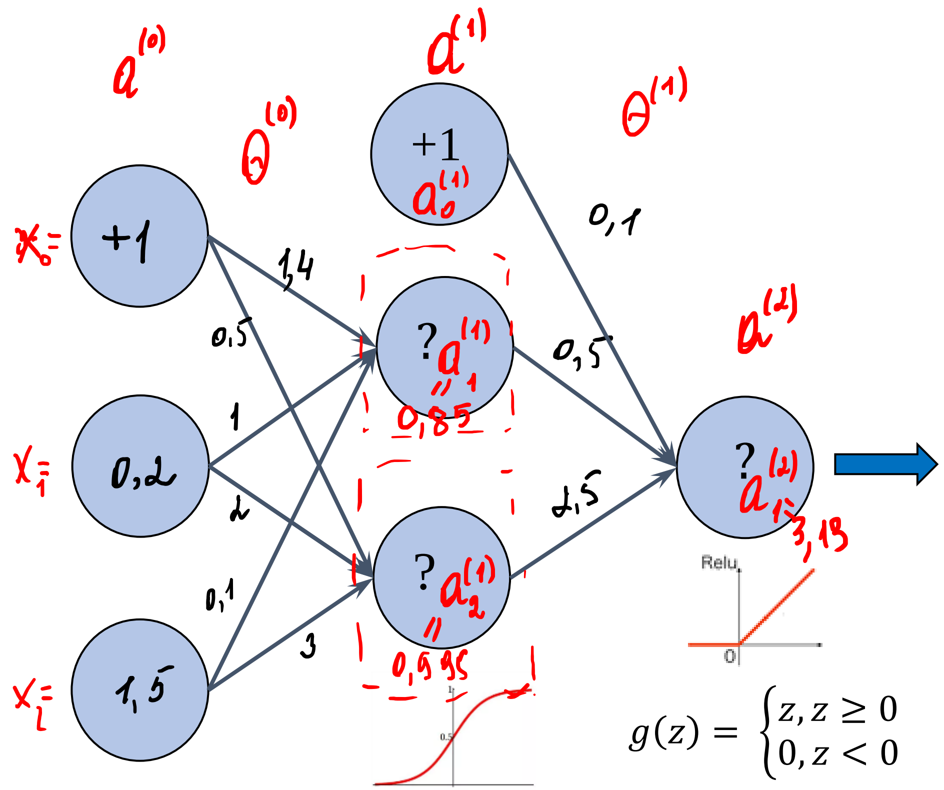


! функция активации м. б. разной на разных слоях

Введение в ИНС

Архитектура сетей прямого распространения / Feed forward neural network

<u>Пример.</u> Вычислить значение функции активации $g(z)$ на последнем слое.



$$g(z) = \begin{cases} z, z \geq 0 \\ 0, z < 0 \end{cases}$$

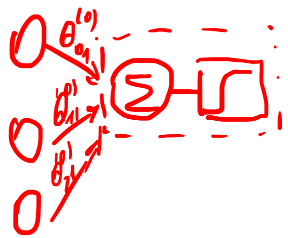
$$f(z) = \frac{1}{1+e^{-z}}$$

$$a_1^{(1)} = sigm(X\overset{\Sigma}{\Theta^{(0)}}) = \frac{1}{1+\exp(-\Theta_0 x_0 - \Theta_1 x_1 - \Theta_2 x_2)} =$$

$$\approx \frac{1}{1+\exp(\underbrace{-1{,}4\cdot 1 - 1\cdot 0{,}2 - 0{,}1\cdot 1{,}5}_{= -1{,}75})} \approx 0{,}85$$

$$a_2^{(1)} = sigm(X\Theta^{(0)}) = (1+\exp(-0{,}5\cdot 1 - \overset{= -5{,}4}{2\cdot 0{,}2 - 3\cdot 1{,}5}))^{-1} \approx 0{,}995$$

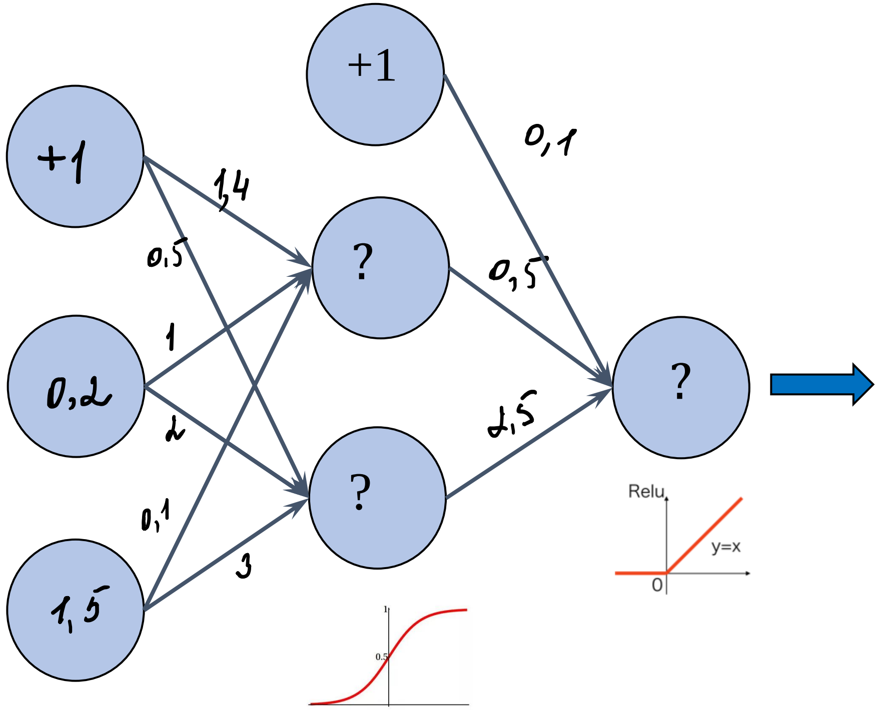
$$a_1^{(2)} = ReLu(X\Theta^{(1)}) = ReLu(0{,}1\cdot 1 + 0{,}5\cdot 0{,}85 + 2{,}5\cdot 0{,}995) = 3{,}13$$



Введение в ИНС

Архитектура сетей прямого распространения / Feed forward neural network

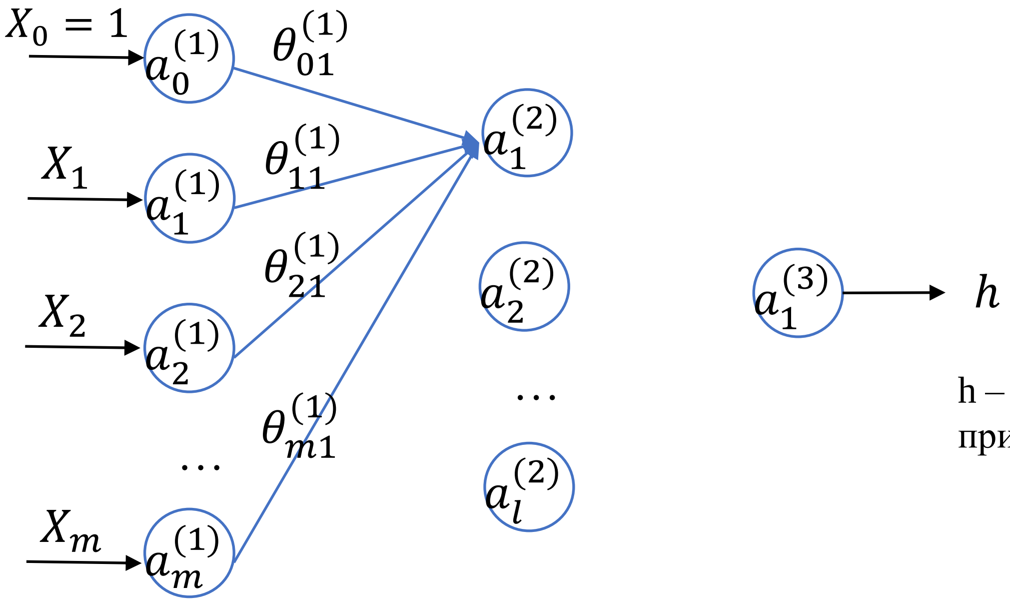
<u>Пример.</u> Запишите вычисления в матричном виде.



Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: Y={0;1}



h – вероятность
принадлежности 1[му] классу

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

p –номер нейрона предыдущего слоя;

q –номер нейрона текущего слоя;

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

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

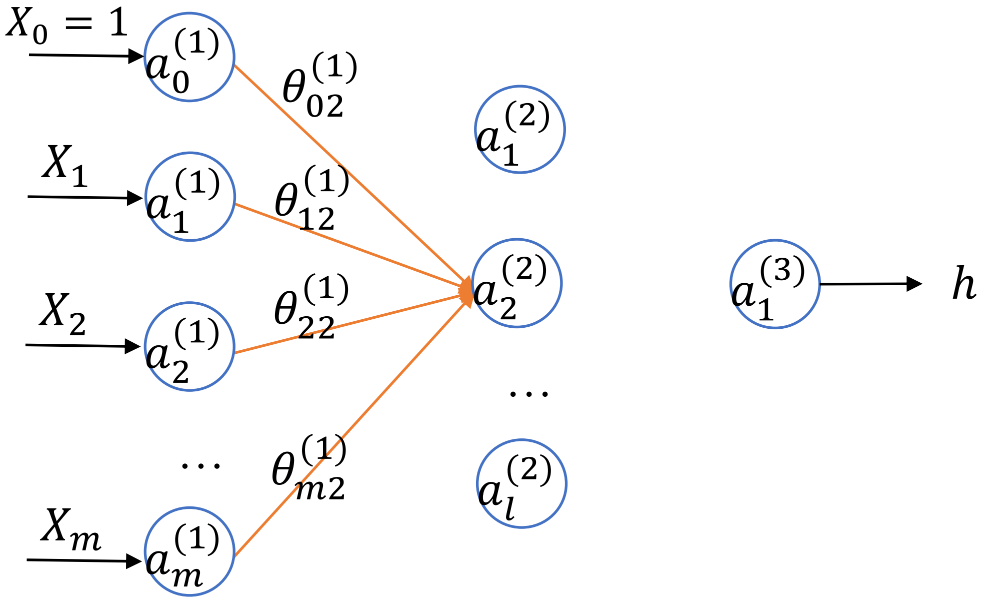
$$Z_1^{(2)} = \theta_{p1}^{(1)} a_p^{(1)}$$

* Используется правило Эйнштейна (сумм. по один. инд.)

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

<u>Два класса: Y={0;1}</u>



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

p –номер нейрона предыдущего слоя;

q –номер нейрона текущего слоя;

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

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

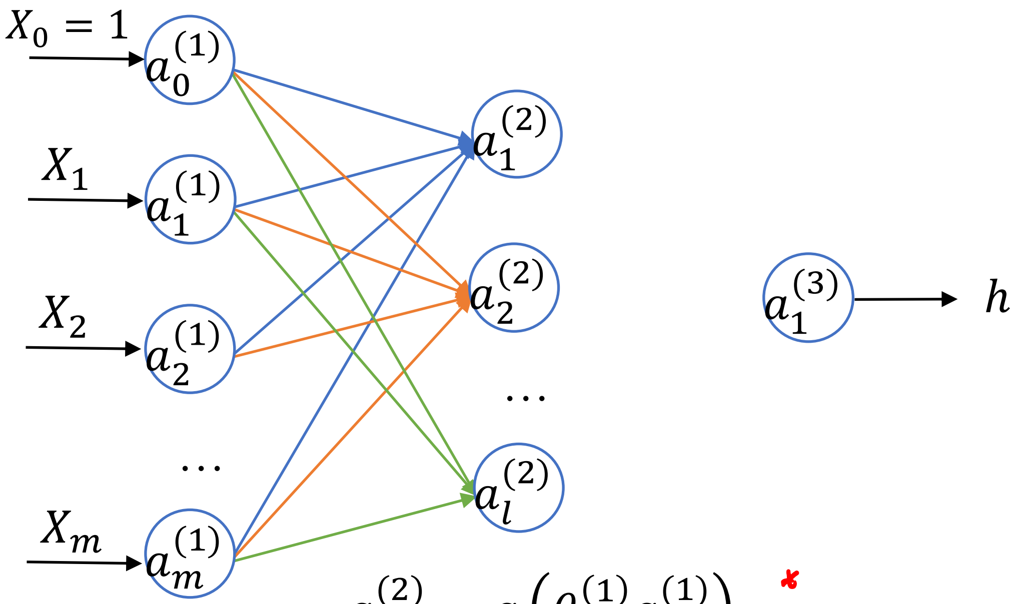
$$Z_2^{(2)} = \theta_{p2}^{(1)}a_p^{(1)}$$

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

<u>Два класса: Y={0;1}.</u>

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



$$\theta^{(1)} = \begin{pmatrix} \theta^{(1)}_{01} & \theta^{(1)}_{02} & & \theta^{(1)}_{0l} \\ \theta^{(1)}_{11} & \theta^{(1)}_{12} & & \theta^{(1)}_{2l} \\ \dots & & & \\ \theta^{(1)}_{m1} & \theta^{(1)}_{m2} & \dots & \theta^{(1)}_{ml} \end{pmatrix}$$

$a_{kq}\ \theta_{lq}$

$$a^{(2)}_q = g\left(\theta^{(1)}_{pq} a^{(1)}_p\right), \quad *$$

* Иссл-е Фурье

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

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

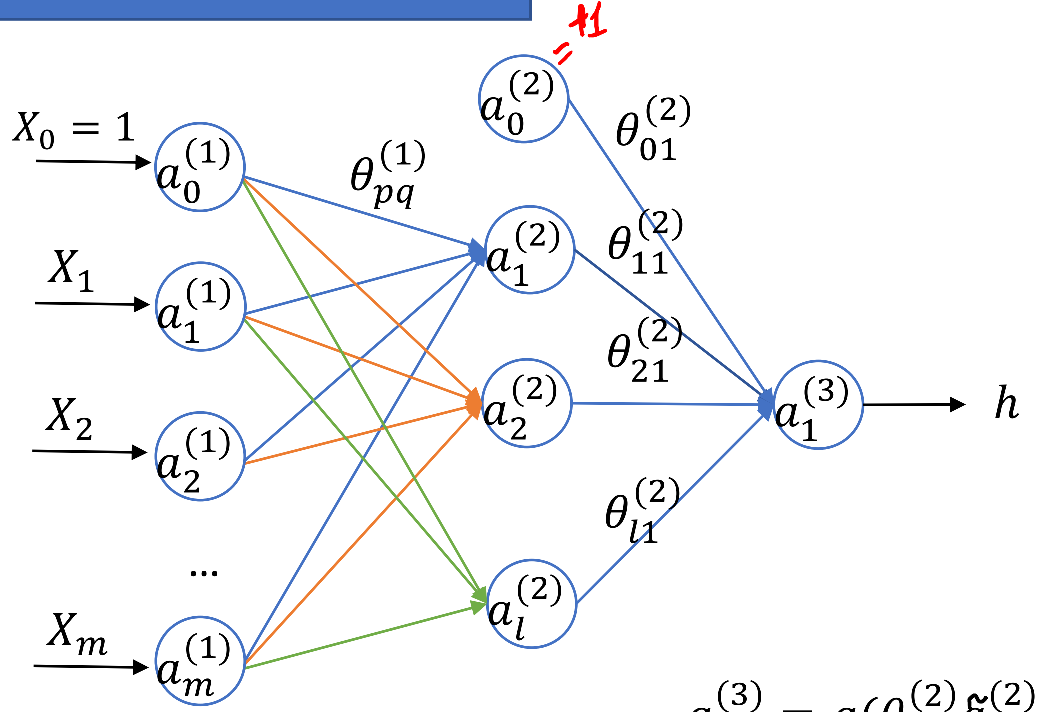
p= $0..m$ −номер нейрона предыдущего слоя;
q= $1..l$ − номер нейрона текущего слоя;
l −кол − во нейронов текущего слоя;

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

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

<u>Два класса: Y={0;1}.</u>



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

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

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

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

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

p – номер нейрона предыдущего слоя;

q – номер нейрона текущего слоя;

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

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

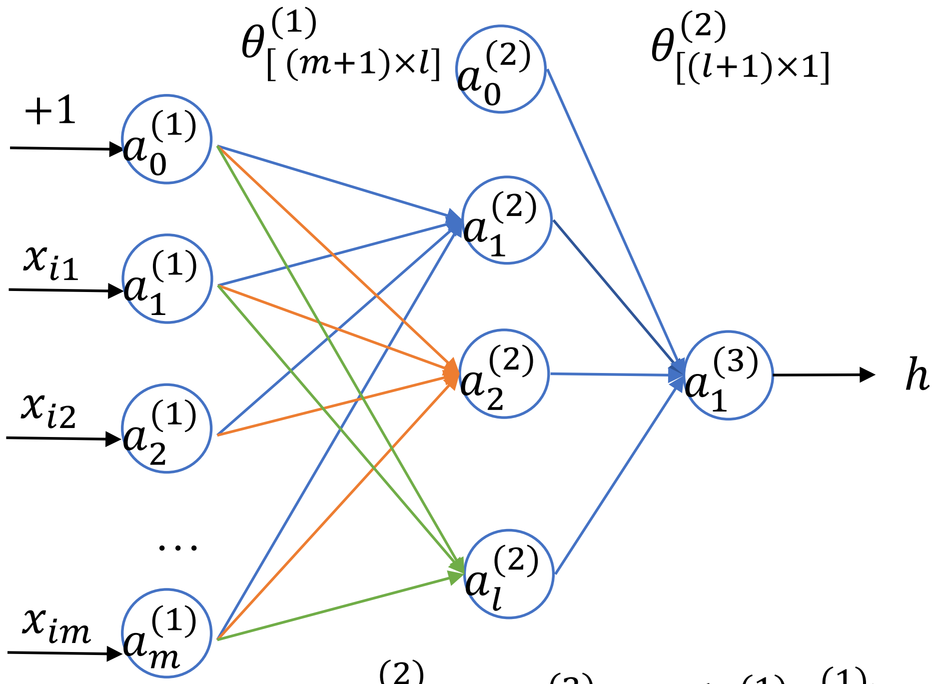
$[n\times 1]$ $[n\times(l+1)]$ $[(l+1)\times 1]$

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: Y={0;1}.

Процедура прямого распространения (Forward propagation). Обобщение формул



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

$$\theta^{(1)} = \begin{pmatrix} \theta^{(1)}_{01} & \theta^{(1)}_{02} & & \theta^{(1)}_{0l} \\ \theta^{(1)}_{11} & \theta^{(1)}_{12} & & \theta^{(1)}_{2l} \\ \dots & & & \\ \theta^{(1)}_{m1} & \theta^{(1)}_{m2} & \dots & \theta^{(1)}_{ml} \end{pmatrix}$$

$$a^{(2)}_{[n\times l]}: \quad a^{(2)} = g(a^{(1)}\theta^{(1)})$$

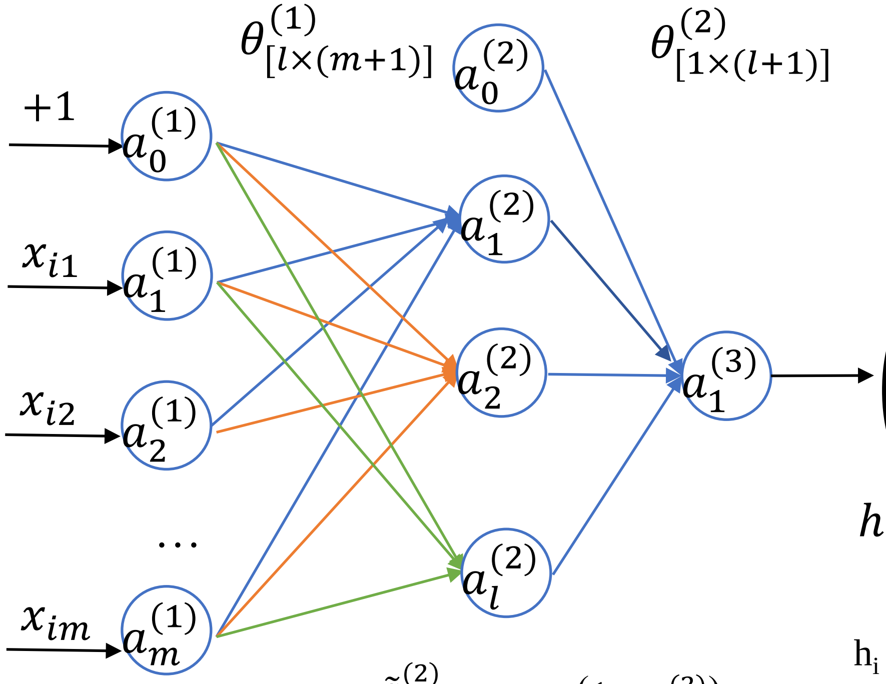
$$\tilde{a}^{(2)} = (1 \quad a^{(2)})$$

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: Y={0;1}.

Процедура прямого распространения (Forward propagation). Обобщение формул



$$\begin{pmatrix} h_1 \\ \dots \\ h_n \end{pmatrix}$$

$$h_{[n\times1]} = a_1^{(3)}$$

$$\tilde{a}^{(2)}_{[n\times(l+1)]} = (1 \quad a^{(2)})$$

$$a^{(3)}_{[n\times1]}\colon\ a^{(3)} = g(\tilde{a}^{(2)}\theta^{(2)})$$

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

h_i – вероятность принадлежности $i^{го}$ объекта классу «1»

Введение в ИНС

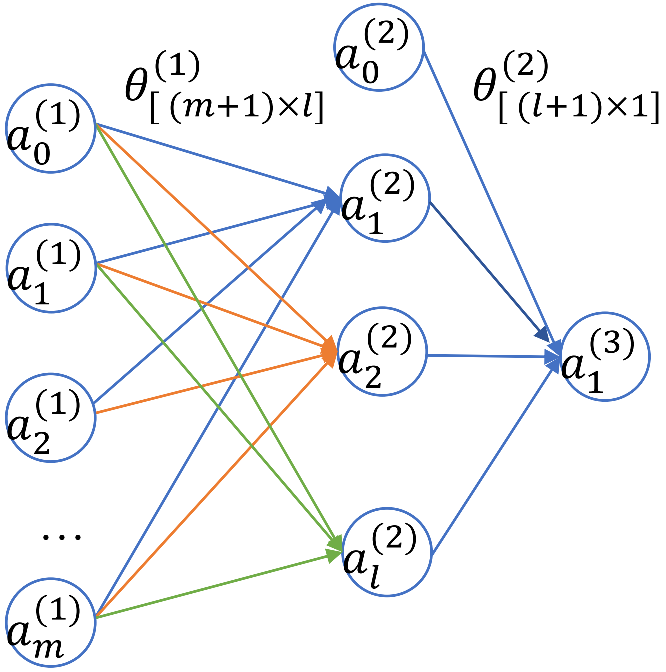
Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: Y={0;1}.

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

Cross Entropy loss

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



Матричный вид:

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

где $I = (1)_{[n\times 1]}$

С учетом регуляризации:

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

Матричный вид:

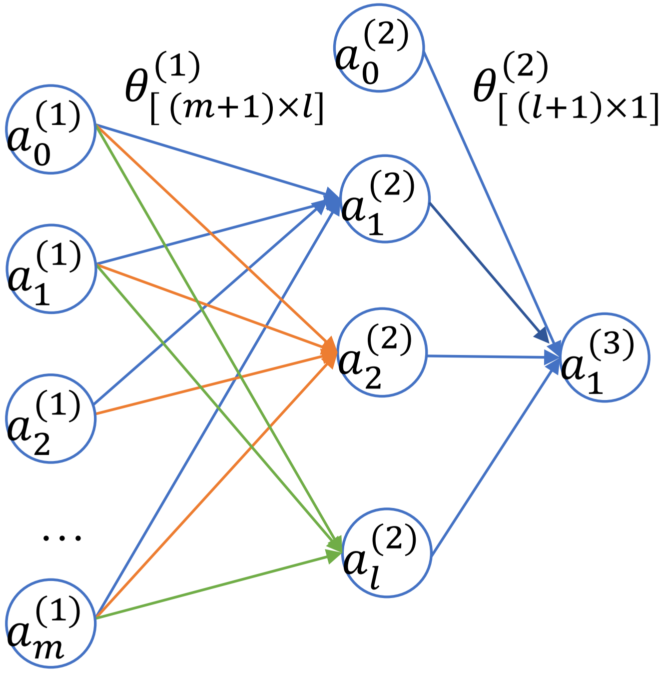
$$J(\theta^{(1)}, \theta^{(2)}) = -\frac{1}{n}(Y^T \ln(h) + (I - Y)^T \ln(I - h)) + \frac{\lambda}{2n}((\theta^{(1)})^T \theta^{(1)} + (\theta^{(2)})^T \theta^{(2)}) \to \min$$

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Два класса: Y={0;1}.

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



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

Процедура прямого распространения ошибки:

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

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

- вычисление значений в каждом нейроне на скрытом слое:

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

- вычисление значений на выходном слое:

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

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

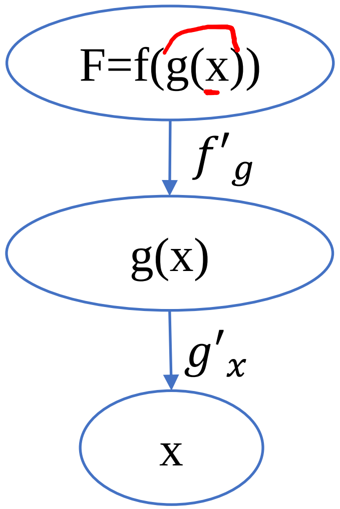
- - вычисление функции ошибки:

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

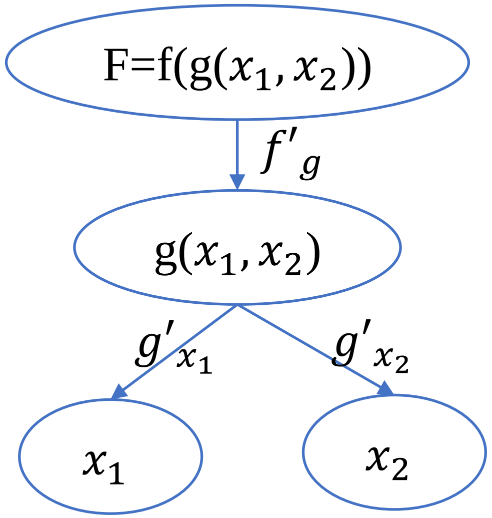
Архитектура сетей прямого распространения. Задача двухклассовой классификации.

<u>Поиск весовых коэффициентов. Метод обратного распространения ошибки.</u>

Правило дифференцирования сложной функции:



$F'_x = ?$



$F'_{x_1} = ?$

$F'_{x_2} = ?$

$F'_x = f'_g * g'_x$

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

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

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

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

$$g(f) = f^{-1}$$

Логистическая функция активации - сигмоида: $\boxed{g(z) = (\underbrace{1 + e^{-z}}_{f(z)})^{-1}}$

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

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

Дифференцирование сигмоиды:

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

$$g'(z) = g'_f \cdot f'_z = -\frac{1}{f^2} \cdot (-e^{-z}) =$$

$$= \frac{(e^{-z} + 1) - 1}{(1+e^{-z})^2} = \frac{1}{1+e^{-z}} - \frac{1}{(1+e^{-z})^2} =$$

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

$$= \frac{1}{1+e^{-z}}\left(1 - \frac{1}{1+e^{-z}}\right) = \mathrm{sigm}(z)(1-\mathrm{sigm}(z))$$

$f(u(x))$

f'_u

$u(x)$

u'_x

x

$$f'_x(u(x)) = f'_u \cdot u'_x$$

$$(e^{Ax})' = e^u \cdot u'_x = e^{Ax} \cdot A$$

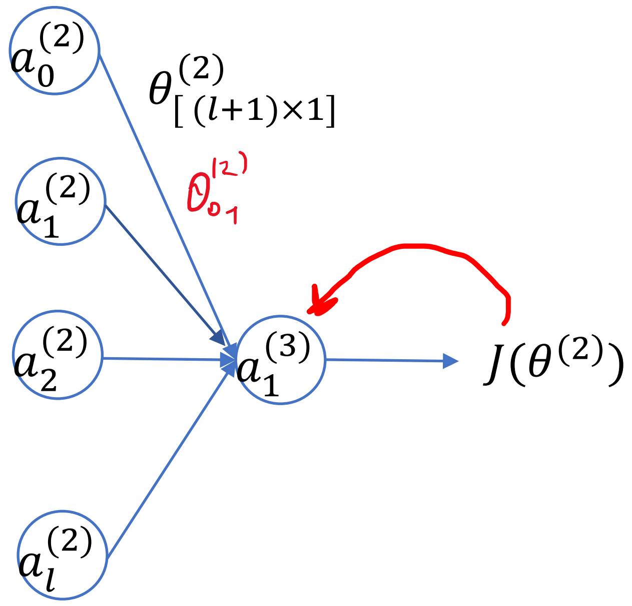
$$f(u) = e^u, \quad u(x) = Ax$$

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

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



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

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

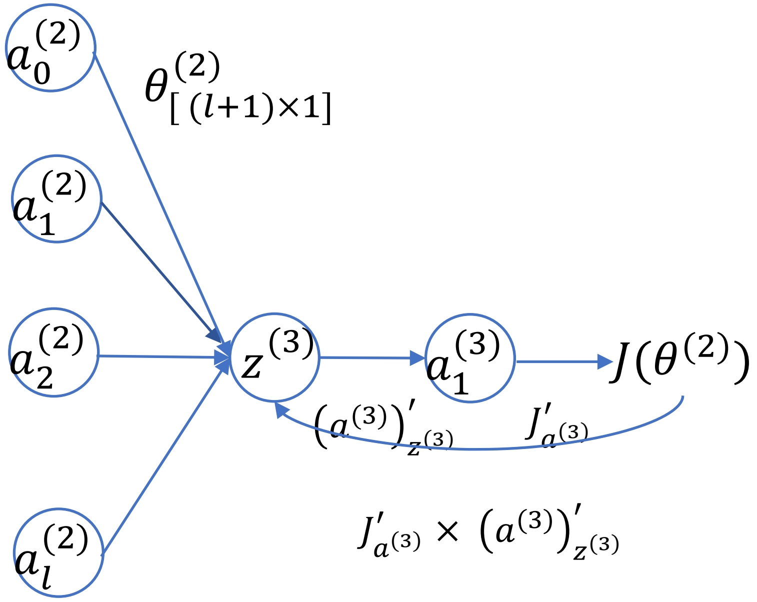
$$y'_{a^{(3)}} = -\frac{1}{n}\left(\frac{y}{a^{(3)}} + \frac{-(1-y)}{1-a^{(3)}}\right) = \frac{1}{n}\cdot\frac{a^{(3)}-y}{a^{(3)}(1-a^{(3)})}$$

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

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.



Дифференцирование J(z):

$$J(z) = J\left(a^{(3)}\left(z^{(3)}\right)\right)$$

$$\left(a^{(3)}\left(z^{(3)}\right)\right)' = \left(1 - a^{(3)}(z)\right) a^{(3)}(z)$$

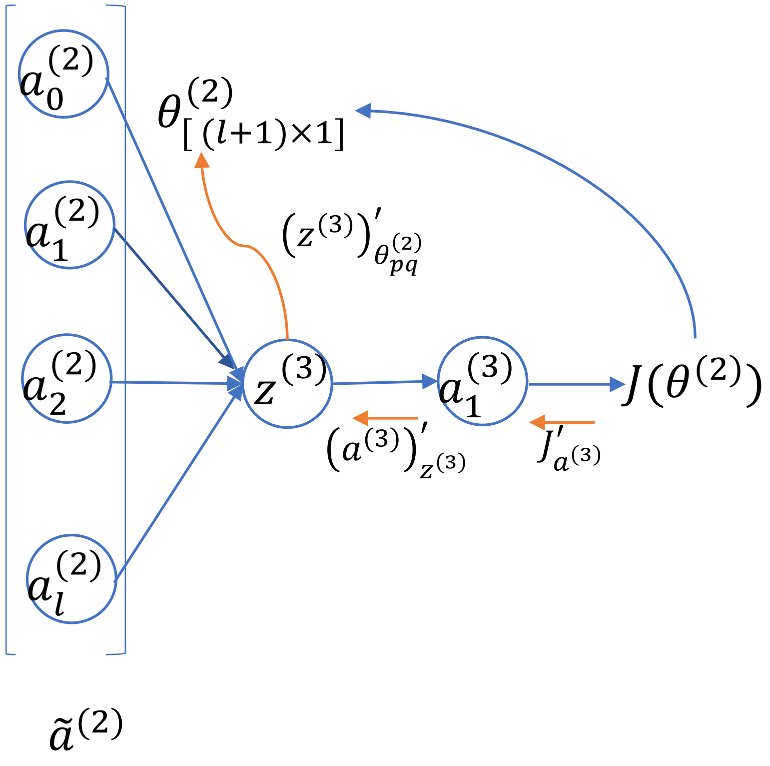
$$J'_{a^{(3)}} = \frac{1}{n}\frac{a^{(3)} - y}{a^{(3)}\left(1 - a^{(3)}\right)}$$

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

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

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



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

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

$$\frac{\partial J}{\partial \theta^{(2)}_{p1}} = \frac{\partial J}{\partial Z^{(3)}}\frac{\partial Z_j^{(3)}}{\partial \theta^{(2)}_{p1}} = \delta^{(3)}\tilde{a}_p^{(2)}, \quad p = 0, \dots l$$

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

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

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

<u>Поиск весовых коэффициентов. Метод обратного распространения ошибки.</u>

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

$$\boxed{a'(z) = a(z)(1-a(z))}$$

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

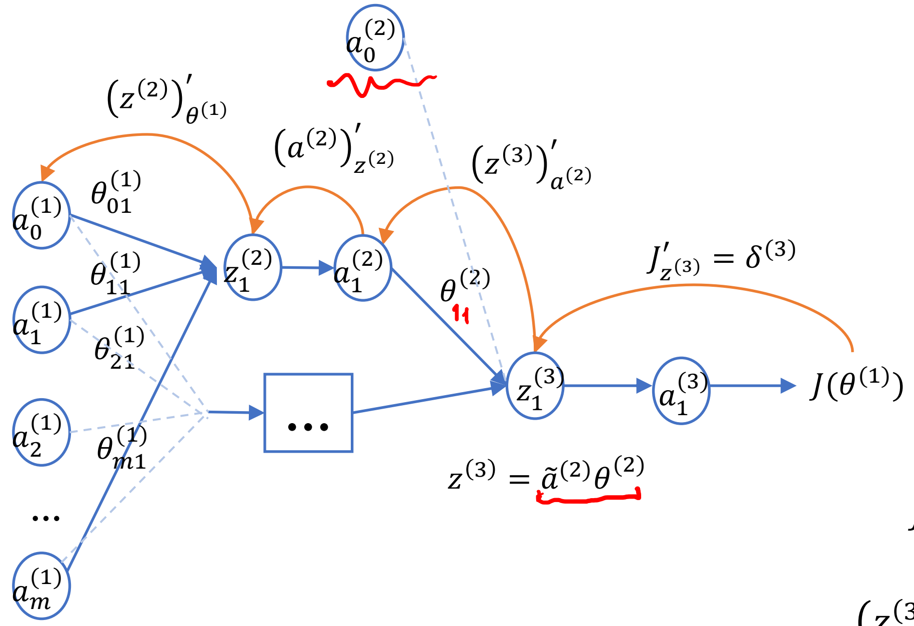


Рис.1. Фрагмент ИНС

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

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

$$J'_{a^{(2)}} = \delta^{(3)}\theta^{(2)}$$

$$(a^{(2)})'_{z^{(2)}} = a^{(2)}(1 - a^{(2)})$$

#без $a_0^{(2)}$

$$J'_{z^{(2)}} = \delta^{(3)}\theta^{(2)}a^{(2)}(1-a^{(2)})$$

$$(z^{(2)})'_{\theta^{(1)}} = a^{(1)}$$

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

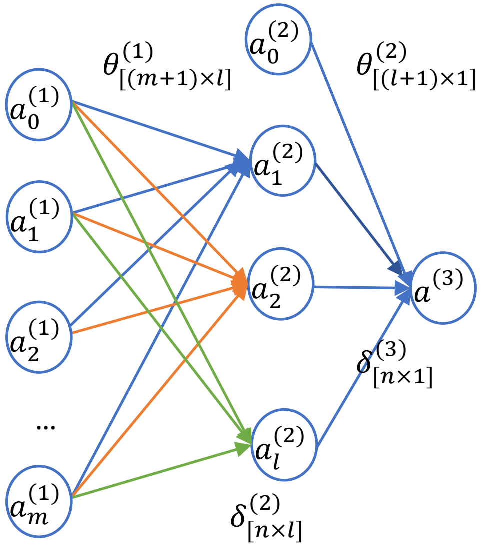
#без $\hat{\delta}_0^{(2)}$

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

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



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

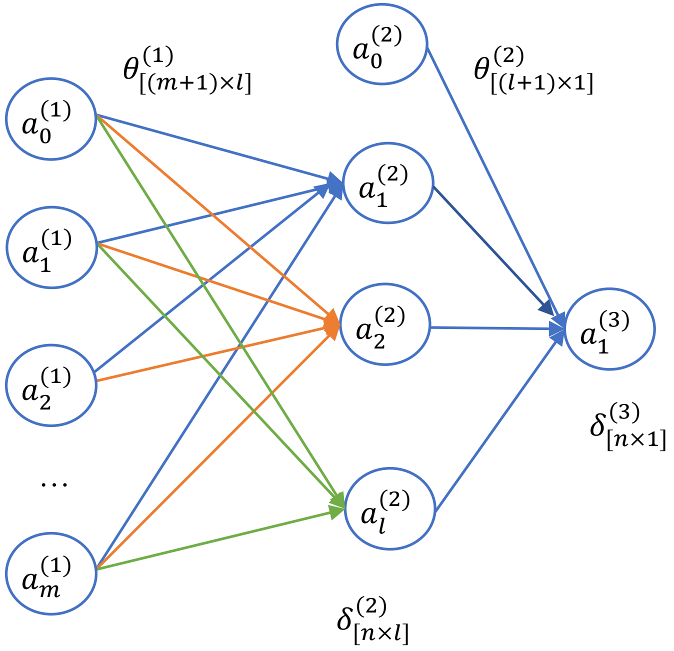
$$\text{где } \delta^{(2)}_{iq} = \hat{\delta}^{(2)}_{iq} a^{(2)}_{iq}\left(1 - a^{(2)}_{iq}\right), \; q = 1..l, \; i=1..n, \; \#\text{без } \hat{\delta}^{(2)}_0$$

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

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.



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

Ошибки на каждом слое:

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

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

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

Компоненты градиента функции ошибки:

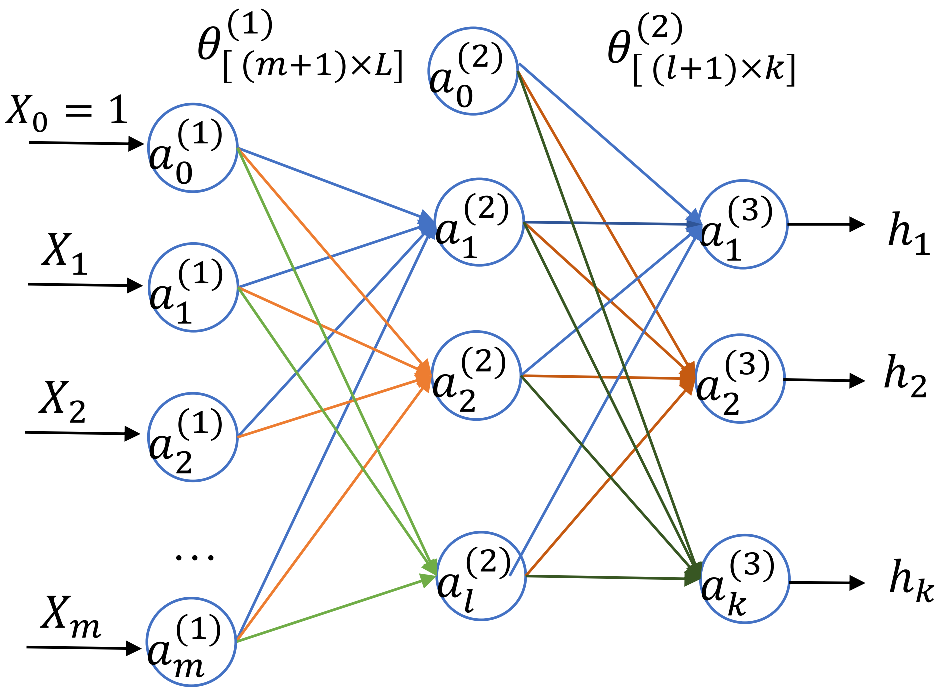
$$\frac{\partial J}{\partial \theta_{p1}^{(2)}} = \delta_i^{(3)} \tilde{a}_{ip}^{(2)},\ p = 0, \dots l,\ i = 1, \dots n$$

$$\frac{\partial J}{\partial \theta_{pq}^{(1)}} = \delta_{iq}^{(2)}\ \tilde{a}_{ip}^{(1)},\quad p{=}0..m,\ q{=}\ 1..l,\ i{=}1,..n \qquad (4)$$

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

Введение в ИНС

Архитектура сетей прямого распространения. Задача многоклассовой классификации.



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

Компоненты градиента функции ошибки:

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

$$\frac{\partial J}{\partial \theta^{(1)}_{pq}} = \delta^{(2)}_{iq}\ \tilde{a}^{(1)}_{ip},\quad p=0..m,\ q= 1..l,\ i=1,..n \qquad (5)$$

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

Компоненты градиента функции ошибки в матричном виде:

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

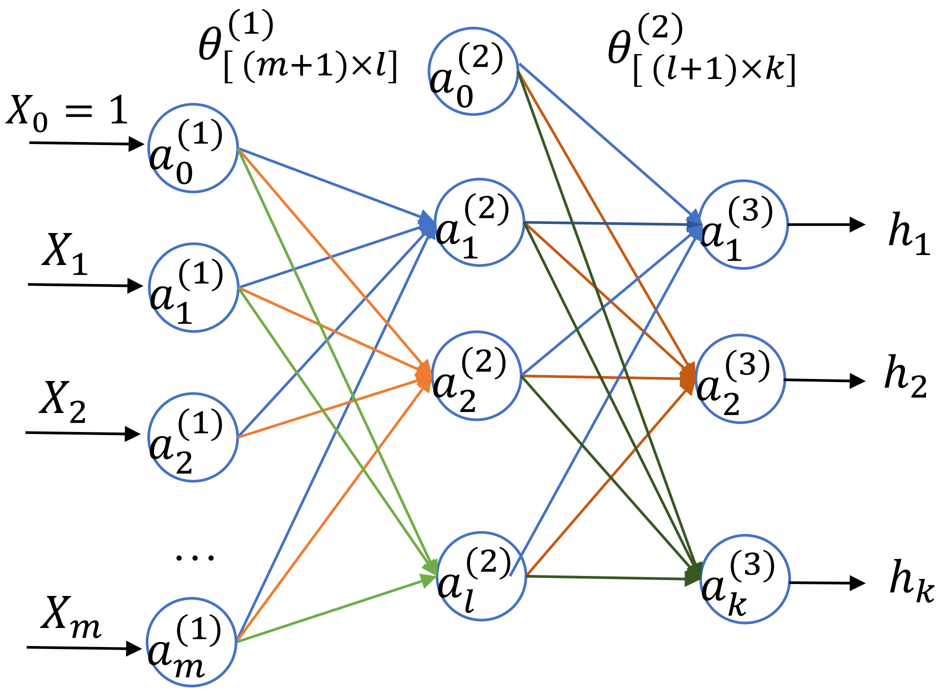
$$\frac{\partial J}{\partial \theta^{(1)}} = \left(\tilde{a}^{(1)}\right)^T \delta^{(2)}, \qquad (5^*)$$

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

$.$ – *означает покомпонентное произведение*

Введение в ИНС

Архитектура сетей прямого распространения. Задача многоклассовой классификации.



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

Обозначение	Описание
$\theta^{(1)}_{[(m+1)\times l]} = (\theta^{(1)}_{pq})$	весовые коэффициенты от 1 слоя ко 2му
$\theta^{(2)}_{[(l+1)\times k]} = (\theta^{(2)}_{pq})$	весовые коэффициенты от 2 слоя к 3му
$a^{(2)}_{[n\times l]} = (a^{(2)}_{ip})$	матрица со значениями в нейронах скрытого слоя для n объектов
$\tilde{a}^{(2)}_{[n\times(l+1)]} = (1 \quad a^{(2)})$	матрица значений в нейронах скрытого слоя, включая bias
$a^{(3)}_{[n\times k]} = (a^{(3)}_{ip})$	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

Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

Компоненты градиента:

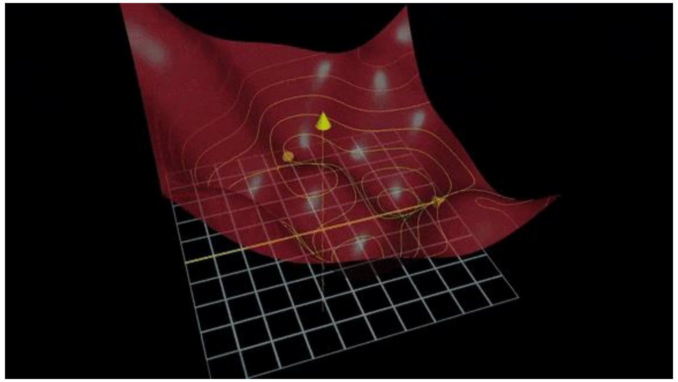
$$\frac{\partial J}{\partial \theta^{(2)}_{01}} = \delta^{(3)}_j \tilde{a}^{(2)}_{jq},\ \text{j}= 1, \dots n,$$

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

$$\frac{\partial J}{\partial \theta^{(1)}_{0q}} = \delta^{(2)}_{pi}\ \tilde{a}^{(1)}_{iq},\quad i\text{=}0..m,\quad //\text{без } \delta^{(2)}_0$$

$$\frac{\partial J}{\partial \theta^{(1)}_{pq}} = \left(\delta^{(2)}_{pi}\ \tilde{a}^{(1)}_{iq} + \frac{\lambda}{n}\theta^{(1)}_{pq}\right),\quad i\text{=}0..m,\ p\text{=}\ 1..l,\ //\text{без } \delta^{(2)}_0$$

$$\Rightarrow \quad \nabla \boldsymbol{J} = \left[\!\left[\begin{matrix} \dfrac{\partial J}{\partial \theta^{(1)}_{pq}} \\ \dfrac{\partial J}{\partial \theta^{(2)}_{pq}} \end{matrix}\right]\!\right]_{[(m+1)\times l+(l+1)\times k]}$$



Gradient descent intuition

Расчет коэффициентов на каждом шаге $s+1$:

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

$\boldsymbol{\Theta}^0 = (\theta^0_{pq})$ – начальное значение коэффициентов;

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

α – шаг градиентного спуска (скорость обучения).

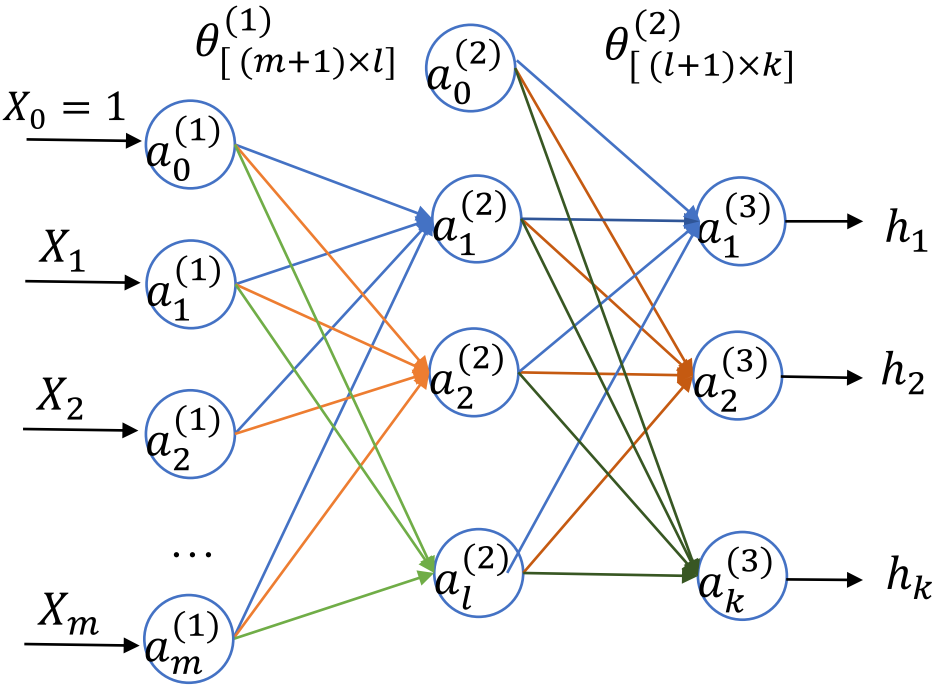
Введение в ИНС

Архитектура сетей прямого распространения. Задача двухклассовой классификации.

Поиск весовых коэффициентов. Метод обратного распространения ошибки.

С учетом регуляризации:

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



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

Компоненты градиента
функции ошибки с учетом регуляризации:

$$\frac{\partial J}{\partial\theta_{0q}^{(2)}} = \delta_{iq}^{(3)}\tilde{a}_{i0}^{(2)},\ q=1,..k,\ i = 1,\dots n$$
$$\frac{\partial J}{\partial\theta_{pq}^{(2)}} = \delta_{iq}^{(3)}\tilde{a}_{ip}^{(2)} + +\frac{\lambda}{n}\theta_{pq}^{(2)},\ p= 1,\dots l,\ q=1,..k,\ i = 1,\dots n$$
$$\frac{\partial J}{\partial\theta_{0q}^{(1)}} = \delta_{iq}^{(2)}\ \tilde{a}_{i0}^{(1)},\quad q= 1..l,\ i=1,..n \qquad (6)$$
$$\frac{\partial J}{\partial\theta_{pq}^{(1)}} = \delta_{iq}^{(2)}\ \tilde{a}_{ip}^{(1)} + \frac{\lambda}{n}\theta_{pq}^{(1)},\ p=1,l,\ q= 1..l,\ i=1,..n$$

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