Постановка задачи

Дано:

$\mathcal{X} = (X_1, \ldots, X_k)$	входные вектора, $X_i \in \mathbb{R}^n$
$\mathcal{A} = (A_1, \ldots, A_k)$	правильные выходные вектора, $A_i \in \mathbb{R}^m$
$(\mathcal{X}, \mathcal{A})$	обучающая выборка
W	вектор весов нейронной сети
$N(W, X)$	функция, соответствующая нейронной сети
$Y = N(W, X)$	ответ нейронной сети, $Y \in \mathbb{R}^m$
$D(Y, A) = \sum_{j=1}^{m}(Y[j] - A[j])^2$	функция ошибки
$D_i(Y) = D(Y, A_i)$	функция ошибки на i-ом примере
$E_i(W) = D_i(N(W, X_i))$	ошибка сети на i-ом примере
$E(W) = \sum_{i=1}^{k} E_i(W)$	ошибка сети на всей обучающей выборке

Найти:

вектор W такой, что $E(W) \to \min$ (обучение на всей выборке)

вектор W такой, что $E_i(W) \to \min$ (обучение на одном примере)

Алгоритм градиентного спуска

1. Инициализировать x_1 случайным значением из $\mathbb{R}$
2. $i := 1$
3. $x_{i+1} = x_i - \varepsilon f'(x_i)$
4. $i + +$
5. $\text{if} \;\; f(x_i) - f(x_{i+1}) > c \;\; \text{goto } 3$

Алгоритм градиентного спуска

1. Инициализировать W_1 случайным значением из $\mathbb{R}^n$
2. $i := 1$
3. $W_{i+1} = W_i - \varepsilon \nabla f(W_i)$
4. $i++$
5. $\text{if} \quad f(W_i) - f(W_{i+1}) > c \quad \text{goto } 3$

Частная производная и градиент

Функция n переменных:

$$f(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$$

Частная производная по i-й переменной:

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) =$$

$$= \lim_{\varepsilon \to 0} \left[f(x_1, x_2, \ldots, x_i + \varepsilon, \ldots, x_n) - f(x_1, x_2, \ldots, x_i, \ldots, x_n) \right] / \varepsilon$$

$$\frac{\partial f}{\partial x_i} :$$

Частная производная и градиент

Функция n переменных:

$$f(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$$

Частная производная по i-й переменной:

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) =$$

$$= \lim_{\varepsilon \to 0} \left[f(x_1, x_2, \ldots, x_i + \varepsilon, \ldots, x_n) - f(x_1, x_2, \ldots, x_i, \ldots, x_n) \right] / \varepsilon$$

$$\frac{\partial f}{\partial x_i} : \mathbb{R}^n \to \mathbb{R}$$

Частная производная и градиент

Функция n переменных:

$$f(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$$

Частная производная по i-й переменной:

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) =$$

$$= \lim_{\varepsilon \to 0} \left[f(x_1, x_2, \ldots, x_i + \varepsilon, \ldots, x_n) - f(x_1, x_2, \ldots, x_i, \ldots, x_n) \right] / \varepsilon$$

$$\frac{\partial f}{\partial x_i} : \mathbb{R}^n \to \mathbb{R}$$

Градиент функции:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

$$\nabla f :$$

Частная производная и градиент

Функция n переменных:

$$f(x_1, \ldots, x_n) : \mathbb{R}^n \to \mathbb{R}$$

Частная производная по i-й переменной:

$$\frac{\partial f}{\partial x_i}(x_1, \ldots, x_n) =$$

$$= \lim_{\varepsilon \to 0} \left[f(x_1, x_2, \ldots, x_i + \varepsilon, \ldots, x_n) - f(x_1, x_2, \ldots, x_i, \ldots, x_n) \right] / \varepsilon$$

$$\frac{\partial f}{\partial x_i} : \mathbb{R}^n \to \mathbb{R}$$

Градиент функции:

$$\nabla f = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \ldots, \frac{\partial f}{\partial x_n} \right)$$

$$\nabla f : \mathbb{R}^n \to \mathbb{R}^n$$

Частные производные

$$f(x, y, z, u) = x^3 + y^u + \sin z^2 u^3$$

$$\frac{\partial f}{\partial x} =$$

$$\frac{\partial f}{\partial y} =$$

$$\frac{\partial f}{\partial z} =$$

Частные производные

$$f(x, y, z, u) = x^3 + y^u + \sin z^2 u^3$$

$$\frac{\partial f}{\partial x} = \quad 3x^2$$

$$\frac{\partial f}{\partial y} =$$

$$\frac{\partial f}{\partial z} =$$

Частные производные

$$f(x, y, z, u) = x^3 + y^u + \sin z^2 u^3$$

$$\frac{\partial f}{\partial x} = 3x^2$$

$$\frac{\partial f}{\partial y} = uy^{u-1}$$

$$\frac{\partial f}{\partial z} =$$

Частные производные

$$f(x, y, z, u) = x^3 + y^u + \sin z^2 u^3$$

$$\begin{aligned}
\frac{\partial f}{\partial x} &= 3x^2 \\
\frac{\partial f}{\partial y} &= uy^{u-1} \\
\frac{\partial f}{\partial z} &= (-\cos z^2 u^3)(u^3 2z)
\end{aligned}$$

Производная сложной функции

$$f = f(x_1, \ldots, x_n)$$

$$x_i = x_i(y_1, \ldots, y_m)$$

$$f(y_1, \ldots, y_m) = f(x_1(y_1, \ldots, y_m), \ldots, x_n(y_1, \ldots, y_m))$$

$$\frac{\partial f}{\partial y_i} = \sum_{j=1}^{n} \frac{\partial f}{\partial x_j} \frac{\partial x_j}{\partial y_i}$$

Производная сложной функции

$$f(x_1, \dots, x_n) = \sum_{k=1}^{n} a_i x_i$$

$$x_i(y_1, \dots, y_m) = \sum_{k=1}^{m} y_k^i$$

$$f(y_1, \dots, y_m) = f(x_1(y_1, \dots, y_m), \dots, x_n(y_1, \dots, y_m))$$

$$\frac{\partial f}{\partial y_j} =$$

Производная сложной функции

$$f(x_1, \ldots, x_n) = \sum_{k=1}^{n} a_i x_i$$

$$x_i(y_1, \ldots, y_m) = \sum_{k=1}^{m} y_k^i$$

$$f(y_1, \ldots, y_m) = f(x_1(y_1, \ldots, y_m), \ldots, x_n(y_1, \ldots, y_m))$$

$$\frac{\partial f}{\partial x_i} = a_i, \ \frac{\partial x_i}{\partial y_j} = i y_j^{i-1}$$

$$\frac{\partial f}{\partial y_j} =$$

Производная сложной функции

$$f(x_1, \ldots, x_n) = \sum_{k=1}^{n} a_i x_i$$

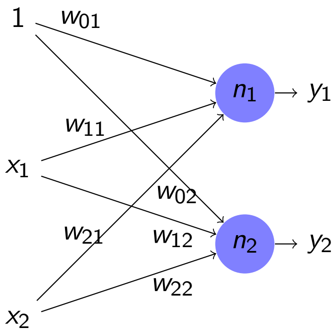
$$x_i(y_1, \ldots, y_m) = \sum_{k=1}^{m} y_k^i$$

$$f(y_1, \ldots, y_m) = f(x_1(y_1, \ldots, y_m), \ldots, x_n(y_1, \ldots, y_m))$$

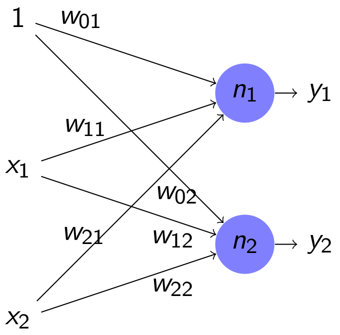
$$\frac{\partial f}{\partial x_i} = a_i, \ \frac{\partial x_i}{\partial y_j} = i y_j^{i-1}$$

$$\frac{\partial f}{\partial y_j} = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial y_j} = \sum_{i=1}^{n} a_i i y_j^{i-1}$$

Обратное распространение ошибки

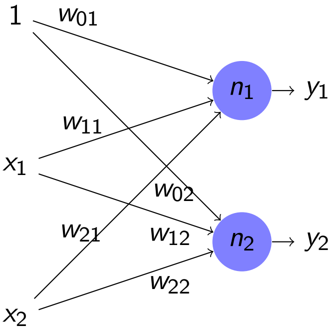


Обратное распространение ошибки



$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

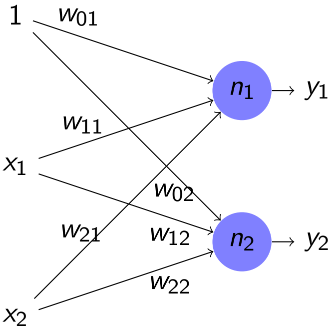
Обратное распространение ошибки



$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} =$$

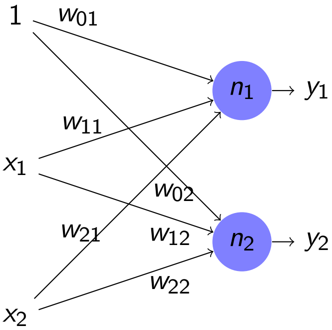
Обратное распространение ошибки



$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1)$$

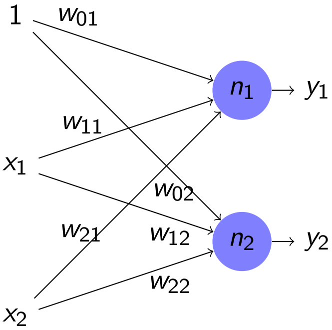
Обратное распространение ошибки



$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

Обратное распространение ошибки

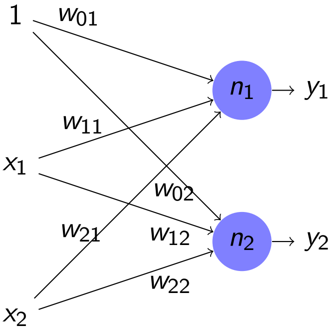


$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_1 = y_1(w_{01}, w_{11}, w_{21}) =$$

Обратное распространение ошибки

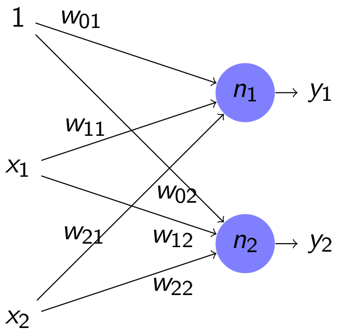


$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_1 = y_1(w_{01}, w_{11}, w_{21}) = f\underbrace{(w_{01} + x_1 w_{11} + x_2 w_{21})}_{S_1}$$

Обратное распространение ошибки



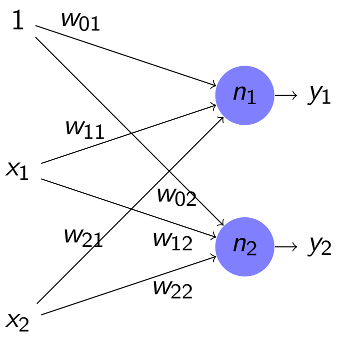
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_1 = y_1(w_{01}, w_{11}, w_{21}) = f\,\underbrace{(w_{01} + x_1 w_{11} + x_2 w_{21})}_{s_1}$$

$$\frac{\partial y_1}{\partial w_{21}} =$$

Обратное распространение ошибки



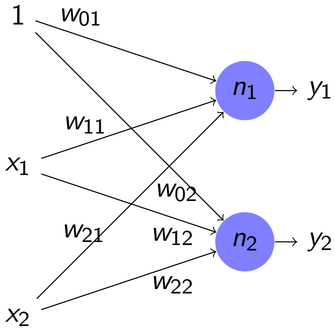
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_1 = y_1(w_{01}, w_{11}, w_{21}) = f\,\underbrace{\left(w_{01} + x_1 w_{11} + x_2 w_{21}\right)}_{S_1}$$

$$\frac{\partial y_1}{\partial w_{21}} = f'(S_1)x_2$$

Обратное распространение ошибки



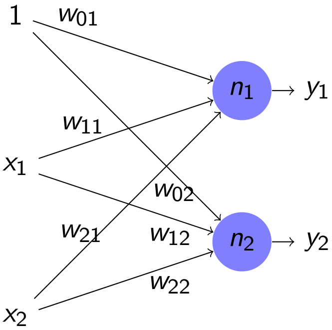
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_2 = y_2(w_{02}, w_{12}, w_{22}) = f\,\underbrace{(w_{02} + x_1 w_{12} + x_2 w_{22})}_{S_2}$$

$$\frac{\partial y_1}{\partial w_{21}} = f'(S_1)x_2$$

Обратное распространение ошибки



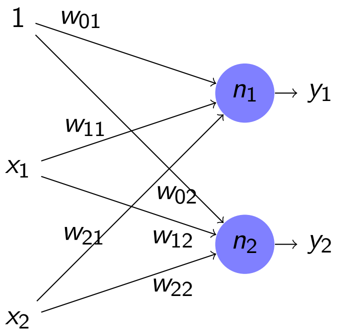
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_2 = y_2(w_{02}, w_{12}, w_{22}) = f\,\underbrace{(w_{02} + x_1 w_{12} + x_2 w_{22})}_{S_2}$$

$$\frac{\partial y_1}{\partial w_{21}} = f'(S_1)x_2 \qquad \frac{\partial y_2}{\partial w_{21}} =$$

Обратное распространение ошибки



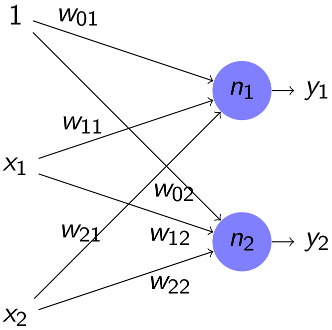
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_2 = y_2(w_{02}, w_{12}, w_{22}) = f\,\underbrace{(w_{02} + x_1 w_{12} + x_2 w_{22})}_{S_2}$$

$$\frac{\partial y_1}{\partial w_{21}} = f'(S_1)x_2 \qquad \frac{\partial y_2}{\partial w_{21}} = 0$$

Обратное распространение ошибки



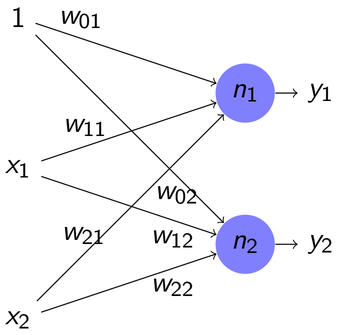
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$\frac{\partial y_1}{\partial w_{21}} = f'(S_1)x_2 \qquad \frac{\partial y_2}{\partial w_{21}} = 0$$

$$E_k(W) = D_k(y_1(w_{01}, w_{11}, w_{21}), y_2(w_{02}, w_{12}, w_{22}))$$

Обратное распространение ошибки



$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

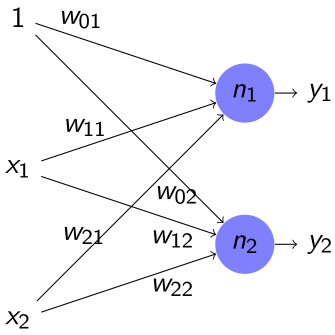
$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$\frac{\partial y_1}{\partial w_{21}} = f'(S_1)x_2 \qquad \frac{\partial y_2}{\partial w_{21}} = 0$$

$$E_k(W) = D_k(y_1(w_{01}, w_{11}, w_{21}), y_2(w_{02}, w_{12}, w_{22}))$$

$$\frac{\partial E_k}{\partial w_{21}} =$$

Обратное распространение ошибки



$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

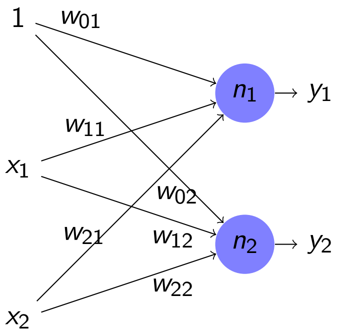
$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$\frac{\partial y_1}{\partial w_{21}} = f'(S_1)x_2 \qquad \frac{\partial y_2}{\partial w_{21}} = 0$$

$$E_k(W) = D_k(y_1(w_{01}, w_{11}, w_{21}), y_2(w_{02}, w_{12}, w_{22}))$$

$$\frac{\partial E_k}{\partial w_{21}} = \frac{\partial D_k}{\partial y_1}\frac{\partial y_1}{\partial w_{21}} + \frac{\partial D_k}{\partial y_2}\frac{\partial y_2}{\partial w_{21}}$$

Обратное распространение ошибки



$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

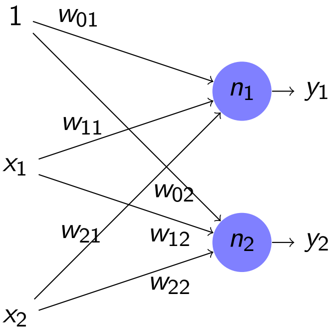
$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$\frac{\partial y_1}{\partial w_{21}} = f'(S_1)x_2 \qquad \frac{\partial y_2}{\partial w_{21}} = 0$$

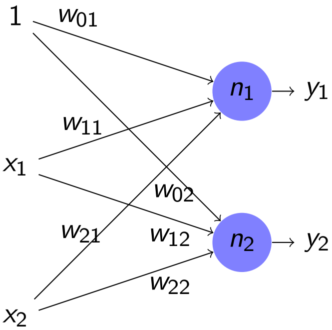
$$E_k(W) = D_k(y_1(w_{01}, w_{11}, w_{21}), y_2(w_{02}, w_{12}, w_{22}))$$

$$\frac{\partial E_k}{\partial w_{21}} = \frac{\partial D_k}{\partial y_1}\frac{\partial y_1}{\partial w_{21}} + \frac{\partial D_k}{\partial y_2}\frac{\partial y_2}{\partial w_{21}} = 2(y_1 - a_1)f'(S_1)x_2$$

Обратное распространение ошибки

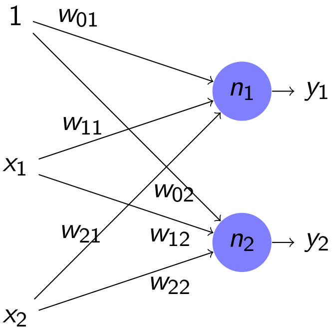


Обратное распространение ошибки



$$D_k(y_1, \ldots, y_n) = (y_1 - a_1)^2 + \ldots + (y_n - a_n)^2$$

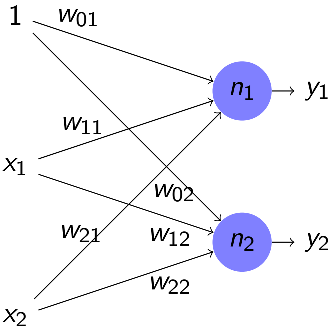
Обратное распространение ошибки



$$D_k(y_1, \ldots, y_n) = (y_1 - a_1)^2 + \ldots + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} =$$

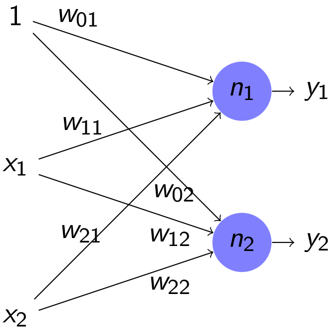
Обратное распространение ошибки



$$D_k(y_1, \dots, y_n) = (y_1 - a_1)^2 + \dots + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

Обратное распространение ошибки

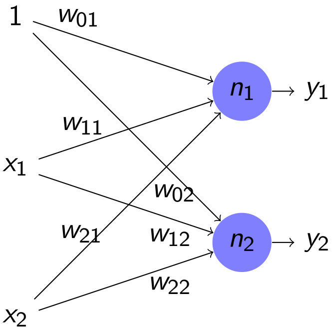


$$D_k(y_1, ..., y_n) = (y_1 - a_1)^2 + ... + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji}$$

Обратное распространение ошибки

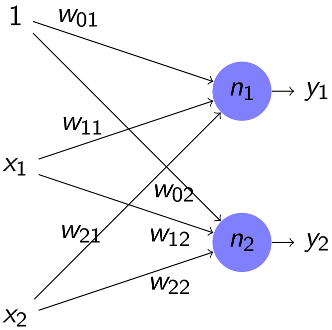


$$D_k(y_1, \ldots, y_n) = (y_1 - a_1)^2 + \ldots + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \textstyle\sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i)$$

Обратное распространение ошибки

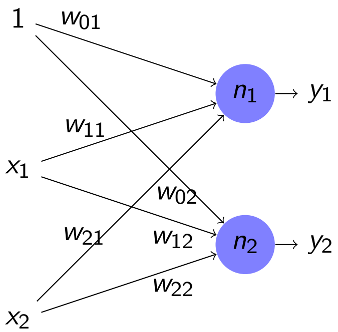


$$D_k(y_1, ..., y_n) = (y_1 - a_1)^2 + ... + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$S_i = \sum_{j=0}^{m} x_j w_{ji}$ $\qquad$ $y_i = f(S_i)$ $\qquad$ $\frac{\partial y_i}{\partial w_{ji}} =$

Обратное распространение ошибки

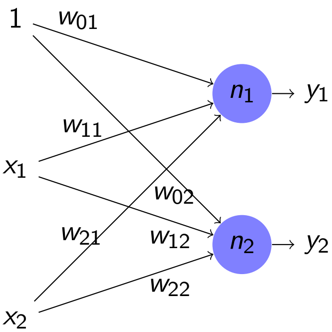


$$D_k(y_1, \ldots, y_n) = (y_1 - a_1)^2 + \ldots + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i) \qquad \frac{\partial y_i}{\partial w_{ji}} = f'(S_i) x_j$$

Обратное распространение ошибки



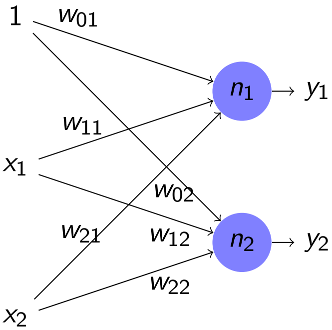
$$D_k(y_1, \ldots, y_n) = (y_1 - a_1)^2 + \ldots + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i) \qquad \frac{\partial y_i}{\partial w_{ji}} = f'(S_i) x_j$$

$$E_k(W) = D_k(y_1(w_{01}, \ldots, w_{mn}), \ldots, y_n(w_{0n}, \ldots, w_{mn}))$$

Обратное распространение ошибки



$$D_k(y_1, ..., y_n) = (y_1 - a_1)^2 + ... + (y_n - a_n)^2$$

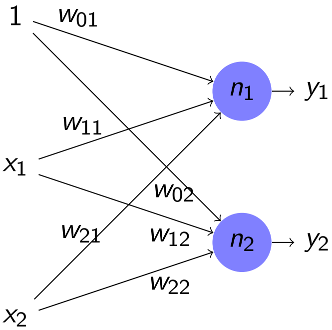
$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i) \qquad \frac{\partial y_i}{\partial w_{ji}} = f'(S_i)x_j$$

$$E_k(W) = D_k(y_1(w_{01}, \ldots, w_{mn}), \ldots, y_n(w_{0n}, \ldots, w_{mn}))$$

$$\frac{\partial E_k}{\partial w_{ji}} =$$

Обратное распространение ошибки



$$D_k(y_1, \ldots, y_n) = (y_1 - a_1)^2 + \ldots + (y_n - a_n)^2$$

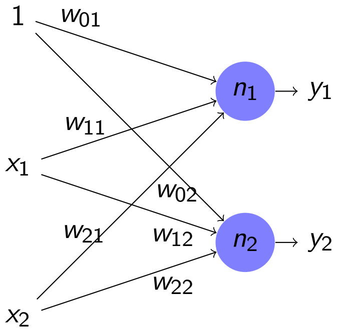
$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i) \qquad \frac{\partial y_i}{\partial w_{ji}} = f'(S_i) x_j$$

$$E_k(W) = D_k(y_1(w_{01}, \ldots, w_{mn}), \ldots, y_n(w_{0n}, \ldots, w_{mn}))$$

$$\frac{\partial E_k}{\partial w_{ji}} = \sum_{l=1}^{n} \frac{\partial D_k}{\partial y_l} \frac{\partial y_l}{\partial w_{ji}}$$

Обратное распространение ошибки



$$D_k(y_1, \ldots, y_n) = (y_1 - a_1)^2 + \ldots + (y_n - a_n)^2$$

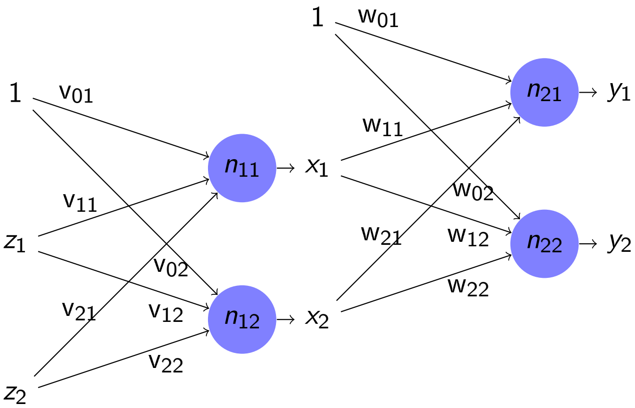
$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i) \qquad \frac{\partial y_i}{\partial w_{ji}} = f'(S_i) x_j$$

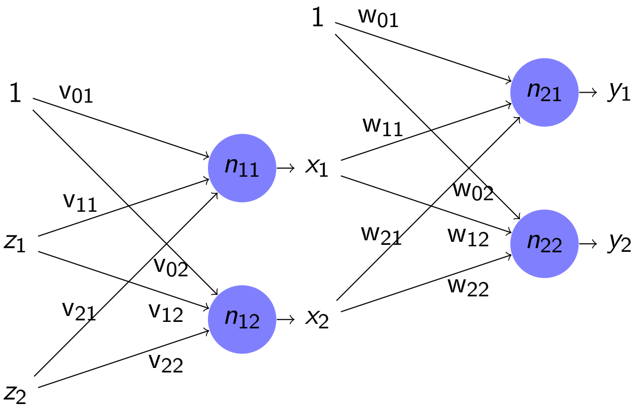
$$E_k(W) = D_k(y_1(w_{01}, \ldots, w_{mn}), \ldots, y_n(w_{0n}, \ldots, w_{mn}))$$

$$\frac{\partial E_k}{\partial w_{ji}} = \sum_{l=1}^{n} \frac{\partial D_k}{\partial y_l} \frac{\partial y_l}{\partial w_{ji}} = 2(y_i - a_i) f'(S_i) x_j$$

Обратное распространение ошибки

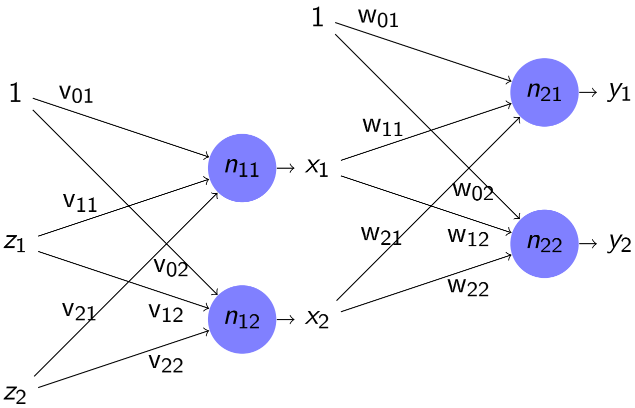


Обратное распространение ошибки



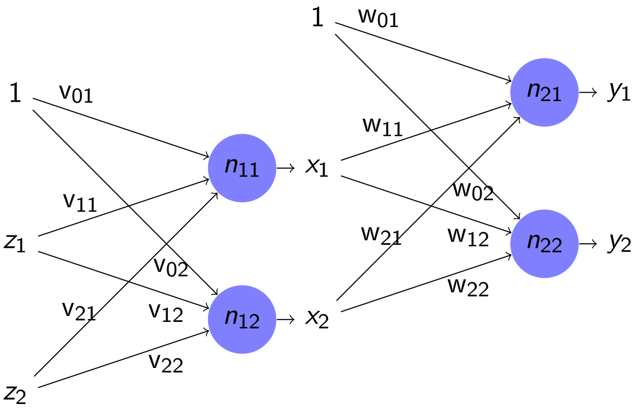
$$E_k(W) = D_k(y_1, \ldots, y_n)$$

Обратное распространение ошибки



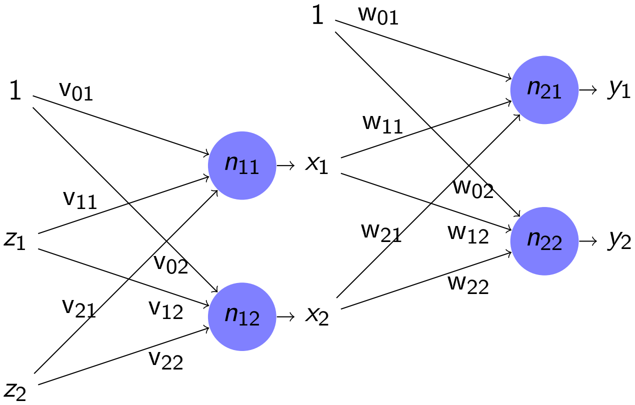
$$E_k(W) = D_k(y_1, \ldots, y_n) \quad y_i = y_i(x_1, \ldots, x_m)$$

Обратное распространение ошибки



$$E_k(W) = D_k(y_1, \ldots, y_n) \quad y_i = y_i(x_1, \ldots, x_m) \quad x_j = x_j(v_{0j}, \ldots, v_{rj})$$

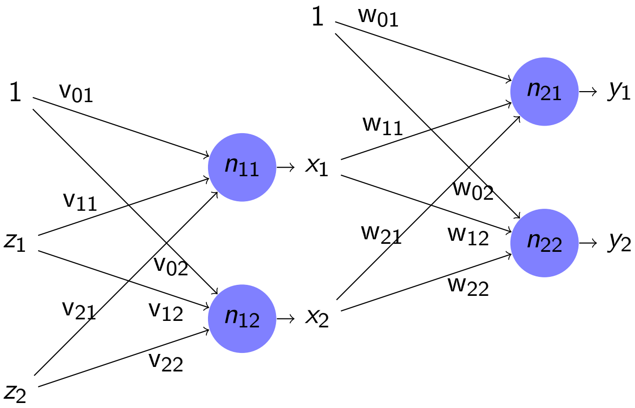
Обратное распространение ошибки



$$E_k(W) = D_k(y_1, \dots, y_n) \quad y_i = y_i(x_1, \dots, x_m) \quad x_j = x_j(v_{0j}, \dots, v_{rj})$$

Если бы $D_k = D_k(x_1, \dots, x_m)$, то $\frac{\partial E_k}{\partial v_{rs}} = \sum_{j=1}^{m} \frac{\partial D_k}{\partial x_j} \frac{\partial x_j}{\partial v_{rs}}$

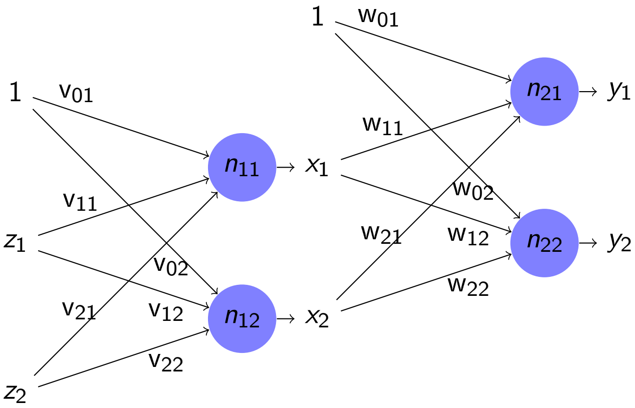
Обратное распространение ошибки



$$E_k(W) = D_k(y_1, \ldots, y_n) \quad y_i = y_i(x_1, \ldots, x_m) \quad x_j = x_j(v_{0j}, \ldots, v_{rj})$$

$$\frac{\partial E_k}{\partial v_{rs}} = \sum_{j=1}^{m} \frac{\partial D_k}{\partial x_j} \frac{\partial x_j}{\partial v_{rs}}$$

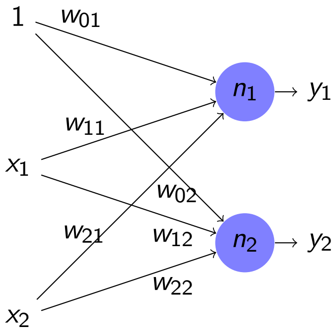
Обратное распространение ошибки



$$E_k(W) = D_k(y_1, \ldots, y_n) \quad y_i = y_i(x_1, \ldots, x_m) \quad x_j = x_j(v_{0j}, \ldots, v_{rj})$$

$$\frac{\partial E_k}{\partial v_{rs}} = \sum_{j=1}^{m} \frac{\partial D_k}{\partial x_j} \frac{\partial x_j}{\partial v_{rs}} \qquad \frac{\partial D_k}{\partial z_l} = \sum_{j=1}^{m} \frac{\partial D_k}{\partial x_j} \frac{\partial x_j}{\partial z_l}$$

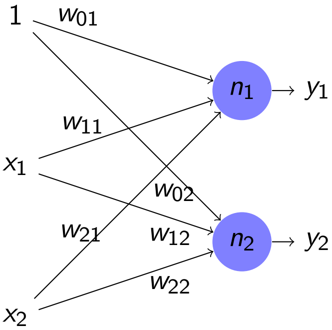
Обратное распространение ошибки



$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \quad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

Обратное распространение ошибки

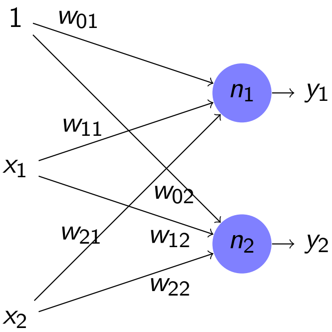


$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \quad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_1 = f\underbrace{\left(w_{01} + x_1 w_{11} + x_2 w_{21}\right)}_{S_1}$$

Обратное распространение ошибки



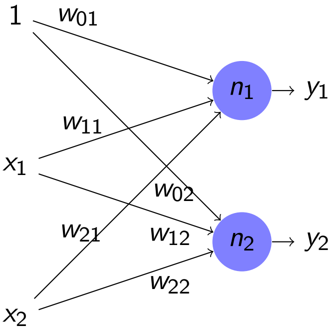
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \quad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_1 = f\underbrace{\left(w_{01} + x_1 w_{11} + x_2 w_{21}\right)}_{s_1}$$

$$\frac{\partial y_1}{\partial x_1} =$$

Обратное распространение ошибки



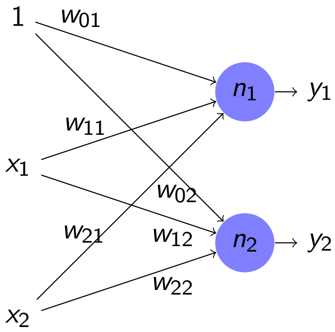
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \quad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_1 = f\,\underbrace{\left(w_{01} + x_1 w_{11} + x_2 w_{21}\right)}_{S_1}$$

$$\frac{\partial y_1}{\partial x_1} = f'(S_1) w_{11}$$

Обратное распространение ошибки



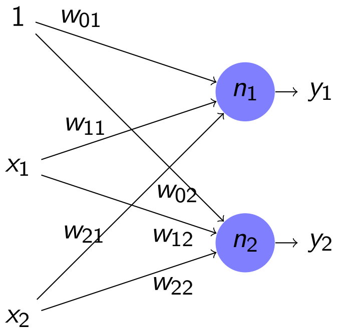
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \quad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_2 = f\,\underbrace{(w_{02} + x_1 w_{12} + x_2 w_{22})}_{S_2}$$

$$\frac{\partial y_1}{\partial x_1} = f'(S_1) w_{11}$$

Обратное распространение ошибки



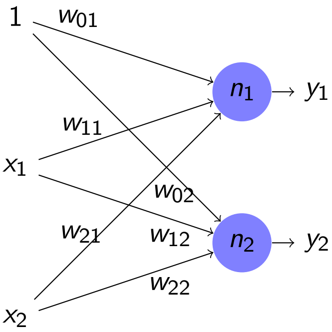
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \quad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_2 = f\underbrace{\left(w_{02} + x_1 w_{12} + x_2 w_{22}\right)}_{S_2}$$

$$\frac{\partial y_1}{\partial x_1} = f'(S_1)w_{11} \qquad \frac{\partial y_2}{\partial x_1} =$$

Обратное распространение ошибки



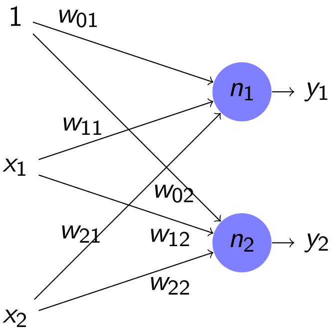
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \quad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$y_2 = f\underbrace{\left(w_{02} + x_1 w_{12} + x_2 w_{22}\right)}_{S_2}$$

$$\frac{\partial y_1}{\partial x_1} = f'(S_1) w_{11} \qquad \frac{\partial y_2}{\partial x_1} = f'(S_2) w_{12}$$

Обратное распространение ошибки



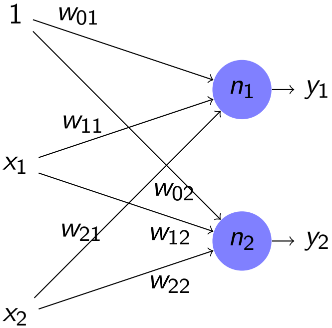
$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \qquad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$\frac{\partial y_1}{\partial x_1} = f'(S_1)w_{11} \qquad \frac{\partial y_2}{\partial x_1} = f'(S_2)w_{12}$$

$$\frac{\partial D_k}{\partial x_1} =$$

Обратное распространение ошибки



$$D_k(y_1, y_2) = (y_1 - a_1)^2 + (y_2 - a_2)^2$$

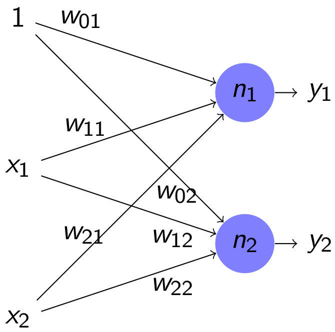
$$\frac{\partial D_k}{\partial y_1} = 2(y_1 - a_1) \quad \frac{\partial D_k}{\partial y_2} = 2(y_2 - a_2)$$

$$\tfrac{\partial y_1}{\partial x_1} = f'(S_1) w_{11} \qquad \tfrac{\partial y_2}{\partial x_1} = f'(S_2) w_{12}$$

$$\frac{\partial D_k}{\partial x_1} = \frac{\partial D_k}{\partial y_1}\frac{\partial y_1}{\partial x_1} + \frac{\partial D_k}{\partial y_2}\frac{\partial y_2}{\partial x_1} =$$

$$= 2(y_1 - a_1) f'(S_1) w_{11} + 2(y_2 - a_2) f'(S_2) w_{12}$$

Обратное распространение ошибки

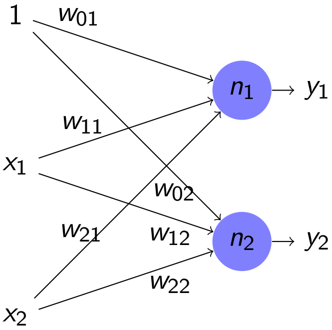


$$D_k(y_1, \dots, y_n) = (y_i - a_i)^2 + \dots + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i) \qquad \frac{\partial y_i}{\partial x_j} =$$

Обратное распространение ошибки

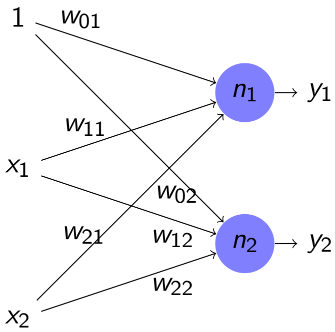


$$D_k(y_1, \dots, y_n) = (y_i - a_i)^2 + \dots + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i) \qquad \frac{\partial y_i}{\partial x_j} = f'(S_i) w_{ji}$$

Обратное распространение ошибки



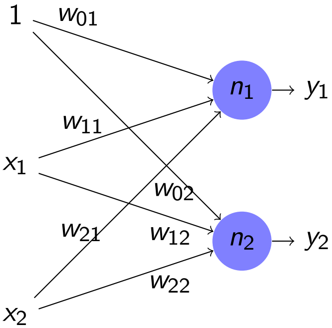
$$D_k(y_1, \ldots, y_n) = (y_i - a_i)^2 + \ldots + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i) \qquad \frac{\partial y_i}{\partial x_j} = f'(S_i) w_{ji}$$

$$\frac{\partial D_k}{\partial x_j} =$$

Обратное распространение ошибки



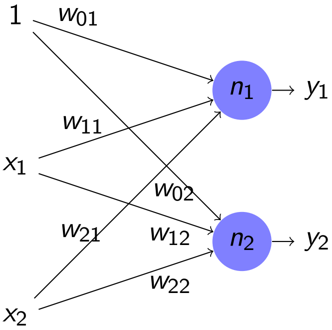
$$D_k(y_1, ..., y_n) = (y_i - a_i)^2 + ... + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i) \qquad \frac{\partial y_i}{\partial x_j} = f'(S_i) w_{ji}$$

$$\frac{\partial D_k}{\partial x_j} = \sum_{i=1}^{n} \frac{\partial D_k}{\partial y_i} \frac{\partial y_i}{\partial x_j} =$$

Обратное распространение ошибки



$$D_k(y_1, \dots, y_n) = (y_i - a_i)^2 + \dots + (y_n - a_n)^2$$

$$\frac{\partial D_k}{\partial y_i} = 2(y_i - a_i)$$

$$S_i = \sum_{j=0}^{m} x_j w_{ji} \qquad y_i = f(S_i) \qquad \frac{\partial y_i}{\partial x_j} = f'(S_i) w_{ji}$$

$$\frac{\partial D_k}{\partial x_j} = \sum_{i=1}^{n} \frac{\partial D_k}{\partial y_i} \frac{\partial y_i}{\partial x_j} =$$

$$= 2 \sum_{i=1}^{n} (y_i - a_i) f'(S_i) w_{ji}$$

Онлайн и оффлайн обучение

Ошибка сети на i-м примере: $E_i(W) = D_i(N(W, X_i))$
Ошибка сети на всей выборке: $E(W) = \sum_{i=1}^{m} E_i(W)$

Онлайн и оффлайн обучение

Ошибка сети на i-м примере: $E_i(W) = D_i(N(W, X_i))$
Ошибка сети на всей выборке: $E(W) = \sum_{i=1}^{m} E_i(W)$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial \sum_{i=1}^{m} E_i(W)}{\partial w_{ij}} = \sum_{i=1}^{m} \frac{\partial E_i}{\partial w_{ij}}$$

Онлайн и оффлайн обучение

Ошибка сети на i-м примере: $E_i(W) = D_i(N(W, X_i))$
Ошибка сети на всей выборке: $E(W) = \sum_{i=1}^{m} E_i(W)$

$$\frac{\partial E}{\partial w_{ij}} = \frac{\partial \sum_{i=1}^{m} E_i(W)}{\partial w_{ij}} = \sum_{i=1}^{m} \frac{\partial E_i}{\partial w_{ij}}$$

$$\nabla E = \sum_{i=1}^{m} \nabla E_i$$

Ускорение обучения

- Правильный выбор коэффициента

$$W_{i+1} = W_i - \varepsilon \nabla f(W_i)$$

(начать с $\varepsilon = 1$)

Ускорение обучения

- Правильный выбор коэффициента

$$W_{i+1} = W_i - \varepsilon \nabla f(W_i)$$

  (начать с $\varepsilon = 1$)
- Повторение алгоритма градиентного спуска на каждом примере (3-5 раз)

Ускорение обучения

- Правильный выбор коэффициента

$$W_{i+1} = W_i - \varepsilon \nabla f(W_i)$$

  (начать с $\varepsilon = 1$)
- Повторение алгоритма градиентного спуска на каждом примере (3-5 раз)
- Правило момента

$$\Delta i = W_i - W_{i-1}$$

$$W_{i+1} = W_i - \varepsilon \nabla f(W_i) + \alpha \Delta_i, \quad \alpha \approx 0.1$$

Ускорение обучения

- Правильный выбор коэффициента

$$W_{i+1} = W_i - \varepsilon \nabla f(W_i)$$

  (начать с $\varepsilon = 1$)
- Повторение алгоритма градиентного спуска на каждом примере (3-5 раз)
- Правило момента

$$\Delta i = W_i - W_{i-1}$$

$$W_{i+1} = W_i - \varepsilon \nabla f(W_i) + \alpha \Delta_i, \quad \alpha \approx 0.1$$

- Стимуляция нейронов

$$W_{i+1} = (1-\gamma) W_i - \varepsilon \nabla f(W_i) + \alpha \Delta_i, \quad \gamma \approx 10^{-4}$$

Ускорение обучения

- Правильный выбор коэффициента

$$W_{i+1} = W_i - \varepsilon \nabla f(W_i)$$

(начать с $\varepsilon = 1$)

- Повторение алгоритма градиентного спуска на каждом примере (3-5 раз)
- Правило момента

$$\Delta i = W_i - W_{i-1}$$

$$W_{i+1} = W_i - \varepsilon \nabla f(W_i) + \alpha \Delta_i, \quad \alpha \approx 0.1$$

- Стимуляция нейронов

$$W_{i+1} = (1-\gamma) W_i - \varepsilon \nabla f(W_i) + \alpha \Delta_i, \quad \gamma \approx 10^{-4}$$

- Адаптивный выбор ε?

Ускорение обучения

- Правильный выбор коэффициента

$$W_{i+1} = W_i - \varepsilon \nabla f(W_i)$$

  (начать с $\varepsilon = 1$)
- Повторение алгоритма градиентного спуска на каждом примере (3-5 раз)
- Правило момента

$$\Delta i = W_i - W_{i-1}$$

$$W_{i+1} = W_i - \varepsilon \nabla f(W_i) + \alpha \Delta_i, \quad \alpha \approx 0.1$$

- Стимуляция нейронов

$$W_{i+1} = (1-\gamma) W_i - \varepsilon \nabla f(W_i) + \alpha \Delta_i, \quad \gamma \approx 10^{-4}$$

- Адаптивный выбор ε?
- Сопряженные градиенты, обучение на гессиане??