

# Лекция 6

## Обучение нейронных сетей

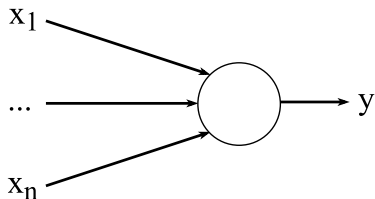
Напоминание: нейронная сеть

# Искусственная нейронная сеть

$$y = h \left( \sum_{i=1}^k w_i x_i + w_0 \right)$$

$w_0, w_1, \dots, w_k$  — параметры нейрона

$h(\cdot)$  — гиперпараметр нейрона



Нейронная сеть = сеть (композиция) нейронов

# Нейронная сеть — пример

$x_1, \dots, x_D$  — ВХОДЫ

$x_0 = 1$  — фиктивный вход

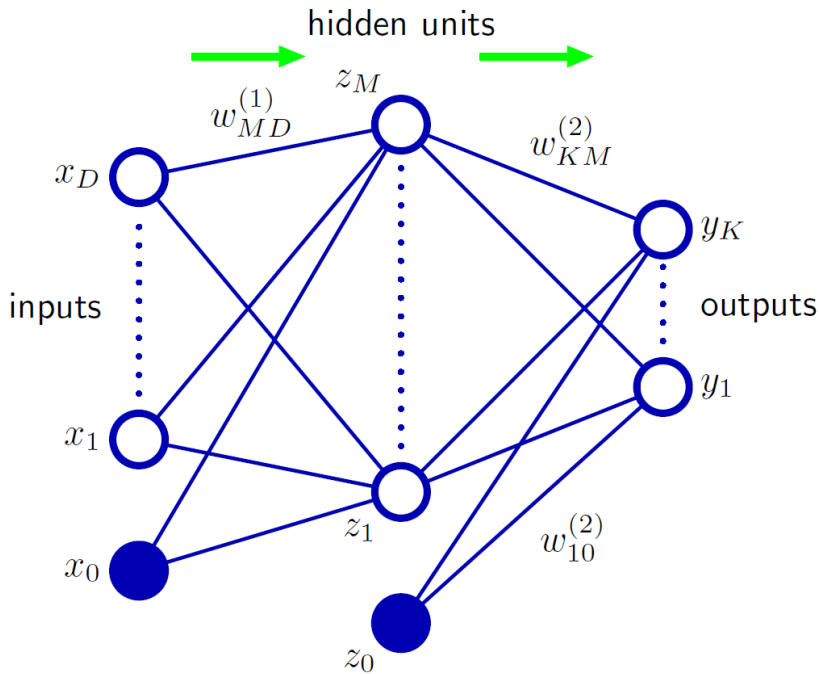
Скрытый слой

$$z_j = h \left( \sum_{i=0}^D w_{ji}^{(1)} x_i \right), \quad j = 1, \dots, M$$

$z_0 = 1$  — фиктивный элемент

Выходной слой

$$y_k = h \left( \sum_{j=0}^M w_{kj}^{(2)} z_j \right), \quad k = 1, \dots, K$$



# Обучение нейронных сетей

# Компоненты обучения

- $\{x_i, y_i\}_{i=1}^N$  — обучающая выборка
- $a(x; w)$  — модель
  - $x$  — входной объект
  - $w$  — параметры (веса входов нейронов)
  - архитектура сети — гиперпараметр
- $L(y, p)$  — функция потерь

$$Q(w) = \sum_{i=1}^N L(y_i, a(x_i; w)) \rightarrow \min_w$$

# Компоненты обучения

- $\{x_i, y_i\}_{i=1}^N$  — обучающая выборка
- $a(x; w)$  — модель
  - $x$  — входной объект
  - $w$  — параметры (веса входов нейронов)
  - архитектура сети — гиперпараметр
- $L(y, p)$  — функция потерь

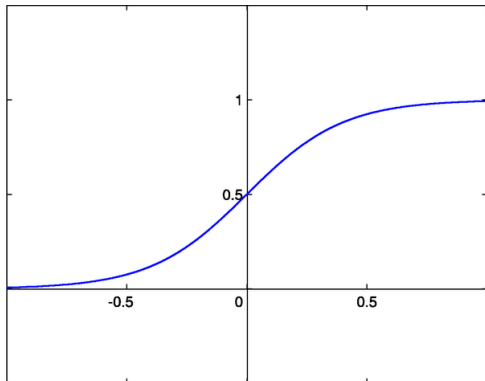
$$Q(w) = \sum_{i=1}^N L(y_i, a(x_i; w)) \rightarrow \min_w$$

*Оптимизация градиентным спуском:  $\frac{\partial Q}{\partial w}$*



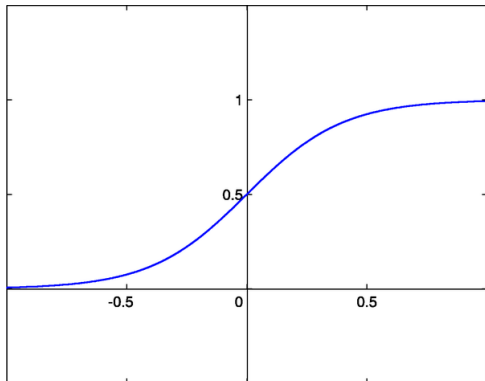
## Напоминание: сигмоида

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



## Напоминание: сигмоида

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



$$\sigma'(z) = \sigma(z)(1 - \sigma(z))$$

## Пример: один нейрон

$$a(x; w) = \sigma \left( \sum_{i=0}^d w_i x^i \right) = \sigma(w^T x)$$

$$L(y, p) = (y - p)^2$$

## Пример: один нейрон

$$a(x; w) = \sigma \left( \sum_{i=0}^d w_i x^i \right) = \sigma(w^T x)$$

$$L(y, p) = (y - p)^2$$

$$\frac{\partial Q(w)}{\partial w} = \sum_{i=1}^N \frac{\partial (y_i - \sigma(w^T x_i))^2}{\partial w} = ?$$

## Пример: один нейрон

$$a(x; w) = \sigma \left( \sum_{i=0}^d w_i x^i \right) = \sigma(w^T x)$$

$$L(y, p) = (y - p)^2$$

$$\frac{\partial Q(w)}{\partial w} = \sum_{i=1}^N \frac{\partial (y_i - \sigma(w^T x_i))^2}{\partial w} = ?$$

Производная композиции функций:

$$\frac{\partial f(g(x))}{\partial x} = \frac{\partial f}{\partial g}(g(x)) \frac{\partial g}{\partial x}(x)$$

## Пример: один нейрон

$$\begin{aligned}\frac{\partial Q(w)}{\partial w} &= \sum_{i=1}^N \frac{\partial (y_i - \sigma(w^T x_i))^2}{\partial w} = \\ &= \sum_{i=1}^N -2(y_i - \sigma(w^T x_i)) \frac{\partial \sigma(w^T x_i)}{\partial w} = \\ &= \sum_{i=1}^N -2(y_i - \sigma(w^T x_i)) \sigma(w^T x_i) (1 - \sigma(w^T x_i)) \frac{\partial (w^T x_i)}{\partial w} = \\ &= \sum_{i=1}^N -2(y_i - \sigma(w^T x_i)) \sigma(w^T x_i) (1 - \sigma(w^T x_i)) x_i\end{aligned}$$

Пример: сеть с одним скрытым слоем

$$z_j(x) = \sigma(w_{1,j}^T x), \quad p(z) = \sigma(w_2^T z)$$

$$L(y, p) = (y - p)^2$$

Пример: сеть с одним скрытым слоем

$$z_j(x) = \sigma(w_{1,j}^T x), \quad p(z) = \sigma(w_2^T z)$$

$$L(y, p) = (y - p)^2$$

$$\frac{\partial Q}{\partial w_2} = \sum_{i=1}^N \frac{\partial L}{\partial p} \frac{\partial p}{\partial w_2}$$



## Пример: сеть с одним скрытым слоем

$$z_j(x) = \sigma(w_{1,j}^T x), \quad p(z) = \sigma(w_2^T z)$$

$$L(y, p) = (y - p)^2$$

$$\frac{\partial Q}{\partial w_2} = \sum_{i=1}^N \frac{\partial L}{\partial p} \frac{\partial p}{\partial w_2}$$

$$\begin{aligned} \frac{\partial Q}{\partial w_{1,j}} &= \sum_{i=1}^N \frac{\partial L}{\partial p} \frac{\partial p}{\partial w_{1,j}} = \sum_{i=1}^N \frac{\partial L}{\partial p} \sum_{s=1}^J \frac{\partial p}{\partial z_s} \frac{\partial z_s}{\partial w_{1,j}} = \\ &= \sum_{i=1}^N \frac{\partial L}{\partial p} \frac{\partial p}{\partial z_j} \frac{\partial z_j}{\partial w_{1,j}} \end{aligned}$$

# Градиент для нейронной сети — общая схема

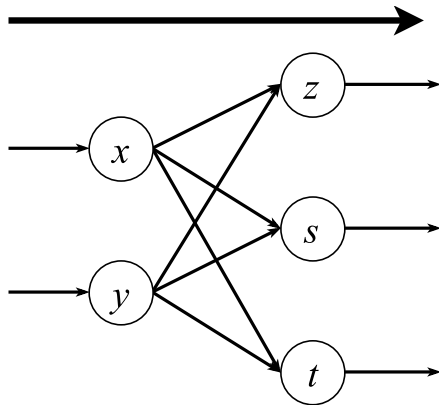
- $a(x; w)$  — сеть  
(композиция дифференцируемых функций)
- $L(y, p)$  — функция потерь  
(дифференцируемая функция)
- $Q(w) = \sum_{i=1}^N L(y_i, a(x_i; w))$  — потери на выборке  
(композиция дифференцируемых функций)

# Градиент для нейронной сети — общая схема

- $a(x; w)$  — сеть  
(композиция дифференцируемых функций)
- $L(y, p)$  — функция потерь  
(дифференцируемая функция)
- $Q(w) = \sum_{i=1}^N L(y_i, a(x_i; w))$  — потери на выборке  
(композиция дифференцируемых функций)

$$\begin{aligned}\frac{\partial Q(\alpha_1, \alpha_2, \dots)}{\partial w_j} &= \sum_s \frac{\partial Q}{\alpha_s} \frac{\partial \alpha_s(\beta_1, \beta_2, \dots)}{\partial w_j} = \\ &= \sum_s \frac{\partial Q}{\alpha_s} \sum_t \frac{\partial \alpha_s}{\partial \beta_t} \frac{\partial \beta_t(\gamma_1, \gamma_2, \dots)}{\partial w_j} = \dots\end{aligned}$$

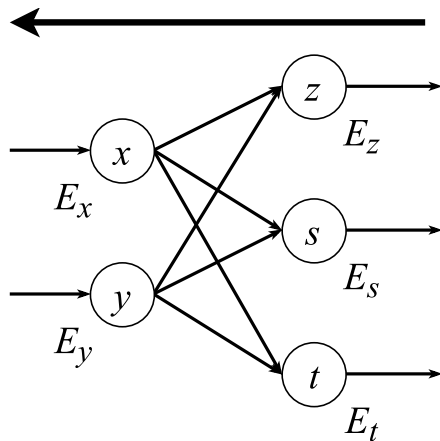
## Прямой ход сети — применение



$$z = h(w_{z,x}x + w_{z,y}y + w_{z,0})$$

# Обратные ход сети — обучение

Backpropagation (обратное распространение ошибки)



$$E_{\alpha} = \frac{\partial Q}{\partial \alpha}$$

$$\alpha = x, y, z, s, t, \dots$$

$$\frac{\partial Q}{\partial w_{\alpha}} = E_{\alpha} \frac{\partial \alpha}{\partial w_{\alpha}}$$

$$E_x = \frac{\partial Q}{\partial x} = \frac{\partial Q}{\partial z} \frac{\partial z}{\partial x} + \frac{\partial Q}{\partial s} \frac{\partial s}{\partial x} + \frac{\partial Q}{\partial t} \frac{\partial t}{\partial x} = E_z \frac{\partial z}{\partial x} + E_s \frac{\partial s}{\partial x} + E_t \frac{\partial t}{\partial x}$$

# Обучение произвольной нейронной сети

- $\{x_i, y_i\}_{i=1}^N$  — обучающая выборка
- $a(x; w)$  — нейронная сеть
- $L(y, p)$  — функция потерь

$$Q(w) = \sum_{i=1}^N L(y_i, a(x_i; w)) \rightarrow \min_w$$

Оптимизация: любой градиентный метод.

# Нейронные сети с регуляризацией

- $\{x_i, y_i\}_{i=1}^N$  — обучающая выборка
- $a(x; w)$  — нейронная сеть
- $L(y, p)$  — функция потерь
- $R(w)$  — функция регуляризации

$$Q(w) = \sum_{i=1}^N L(y_i, a(x_i; w)) + R(w) \rightarrow \min_w$$

Оптимизация: любой градиентный метод.

# Источники

- Bishop C. Pattern Recognition and Machine Learning.  
Глава 5.