

Перцептрон и его друзья

Хахулин Тарас

МФТИ
ФРТК

Deep Learning School, 2017

Outline

1 Перцептрон

История развития

Перцептрон: введение

Логика на перцептроне

XOR проблема

2 Обучение перцептрона

Многослойные сети

Градиентный спуск

3 Различные активационные функции

Друзья перцептрона

4 Обучение v.2

Правила обновления весов

Стохастический градиентный спуск

План лекции

1 Перцептрон

История развития

Перцептрон: введение

Логика на перцептроне

XOR проблема

2 Обучение перцептрона

Многослойные сети

Градиентный спуск

3 Различные активационные функции

Друзья перцептрона

4 Обучение v.2

Правила обновления весов

Стохастический градиентный спуск

История развития

Активационная функция:

- Модель Розенблата.
- Критика со всех сторон.
- Застой на долгие годы.



План лекции

1 Перцептрон

История развития

Перцептрон: введение

Логика на перцептроне

XOR проблема

2 Обучение перцептрона

Многослойные сети

Градиентный спуск

3 Различные активационные функции

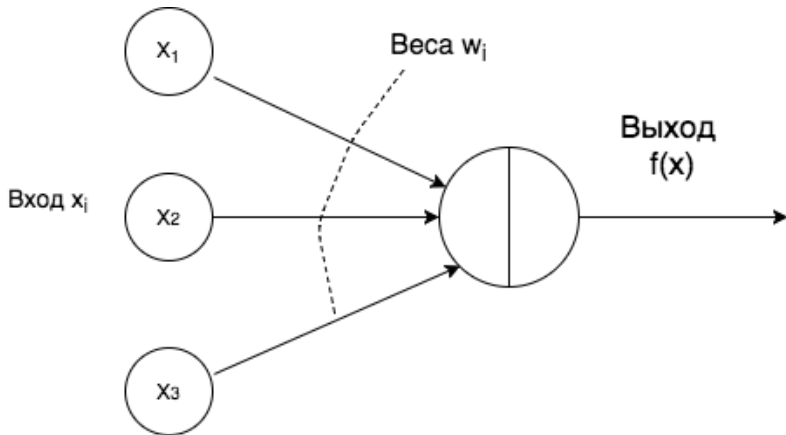
Друзья перцептрона

4 Обучение v.2

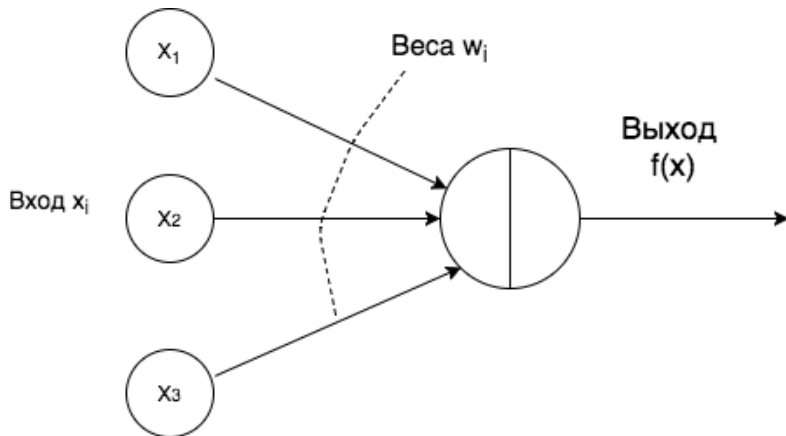
Правила обновления весов

Стохастический градиентный спуск

Модель перцептрона



Модель перцептрона



**I SURVIVED THE
A. I. WINTER**



(1974-80)

Модель перцептрона

$$f(x, w, b) = \begin{cases} 1 & \text{если } \sum_{i=1}^n x_i w_i + b > 0 \\ 0 & \text{если } \sum_{i=1}^n x_i w_i + b \leq 0 \end{cases}$$

Где x_i - входы $i=\{1, n\}$, w_i - веса модели,
 b - смещение, $f(x)$ - выходная активация

План лекции

1 Перцептрон

История развития

Перцептрон: введение

Логика на перцептроне

XOR проблема

2 Обучение перцептрона

Многослойные сети

Градиентный спуск

3 Различные активационные функции

Друзья перцептрона

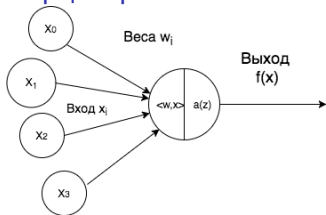
4 Обучение v.2

Правила обновления весов

Стохастический градиентный спуск

Модель перцептрона

Перцептрон



$$f(x, w, b) = \begin{cases} 1 & \text{если } \langle w, x \rangle > 0 \\ 0 & \text{если } \langle w, x \rangle \leq 0 \end{cases}$$

Упражнение

Попробуйте понять какую функцию воспроизводит вектор

$$\vec{w} = \begin{bmatrix} -1, 5 \\ 1 \\ 1 \end{bmatrix} \quad (1)$$



**Logic: another thing that
penguins aren't very good at.**

План лекции

1 Перцептрон

История развития

Перцептрон: введение

Логика на перцептроне

XOR проблема

2 Обучение перцептрона

Многослойные сети

Градиентный спуск

3 Различные активационные функции

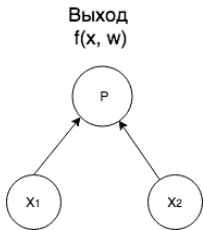
Друзья перцептрона

4 Обучение v.2

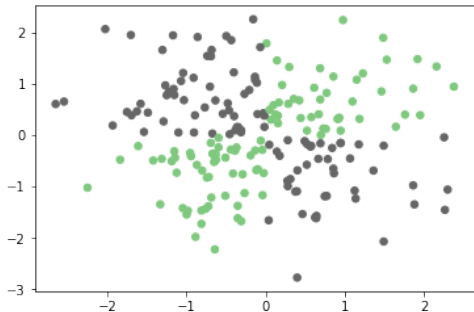
Правила обновления весов

Стохастический градиентный спуск

XOR проблема

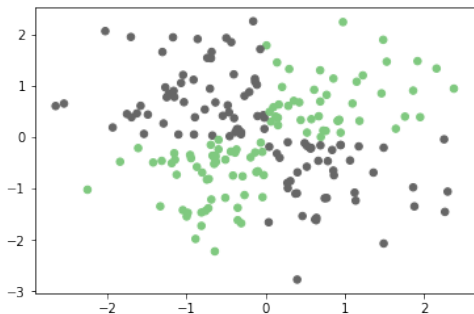
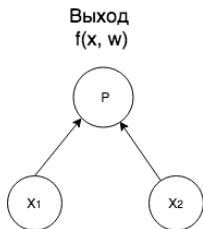


$$\begin{cases} w_1 + w_2 > P \\ w_1 < P \\ w_2 < P \\ 0 > P \end{cases}$$



Кто видит
противоречие?

XOR проблема



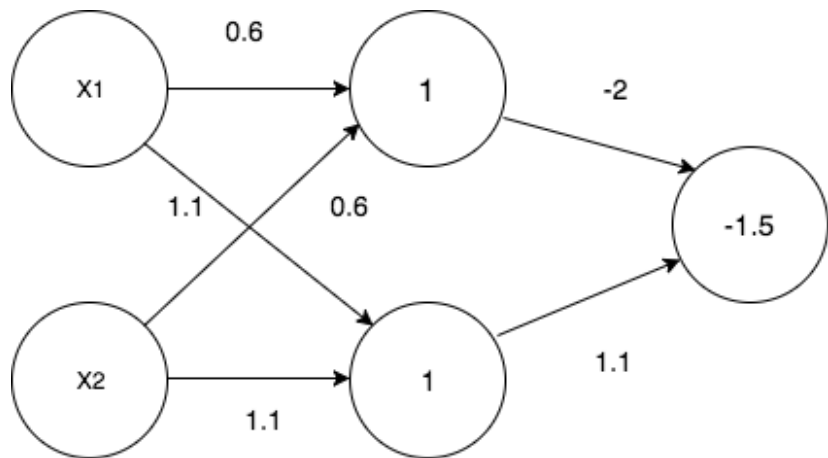
А решение?

A close-up of Leonardo DiCaprio from the movie Inception, looking slightly to the right with a serious expression. The lighting is warm and dramatic, with a blurred background showing architectural elements.

PERCEPTRONS ARE TOO LINEAR

WE MUST GO DEEPER

XOR решение



План лекции

1 Перцептрон

История развития

Перцептрон: введение

Логика на перцептроне

XOR проблема

2 Обучение перцептрона

Многослойные сети

Градиентный спуск

3 Различные активационные функции

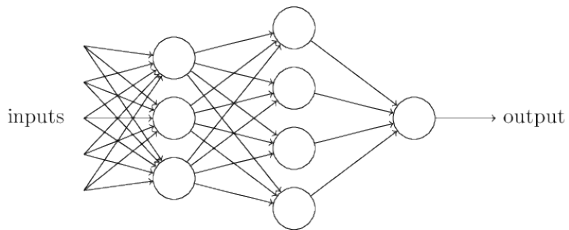
Друзья перцептрона

4 Обучение v.2

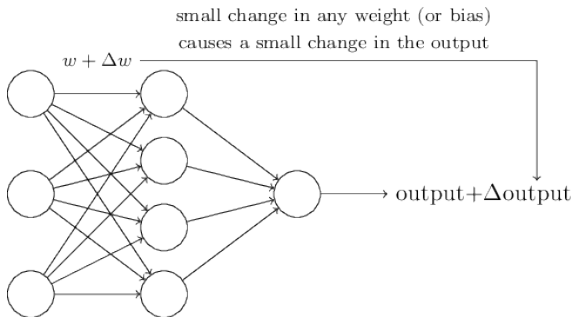
Правила обновления весов

Стохастический градиентный спуск

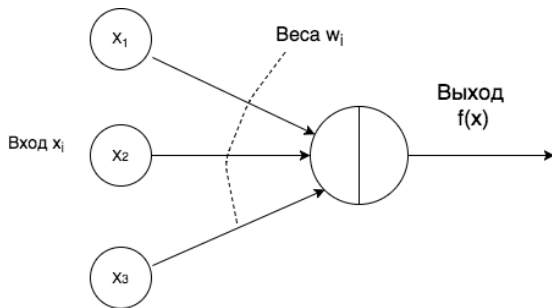
Пример



Обучение



Обучение



Подвинь веса при обучении:

$$\Delta w_i = (\hat{y} - y)x_i, \hat{y} = f(x, w)$$

План лекции

1 Перцептрон

История развития

Перцептрон: введение

Логика на перцептроне

XOR проблема

2 Обучение перцептрона

Многослойные сети

Градиентный спуск

3 Различные активационные функции

Друзья перцептрона

4 Обучение v.2

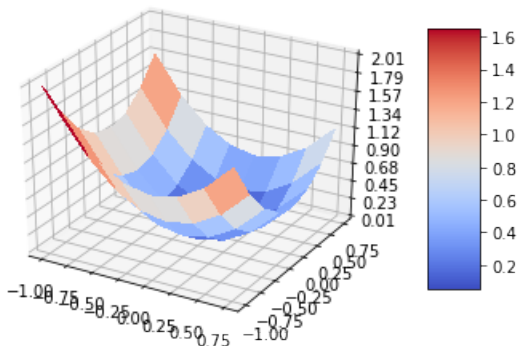
Правила обновления весов

Стохастический градиентный спуск

Постановка

Задача: найти минимум функции $J(\vec{w}) = \frac{1}{2n} \sum_i (f(\langle w, x \rangle) - y)^2$

Варианты плоскостей:



Градиентный спуск

- Задача: найти минимум функции $J(\vec{w})$
- Инициализируем \vec{w}_0

Градиентный спуск

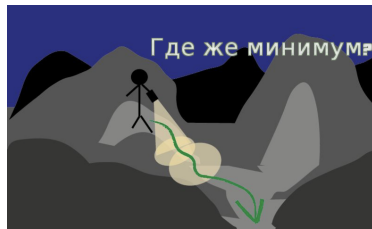
- Задача: найти минимум функции $J(\vec{w})$
- Инициализируем \vec{w}_0
- Посчитаем градиент. $\nabla J(\vec{w}_0)$

Градиентный спуск

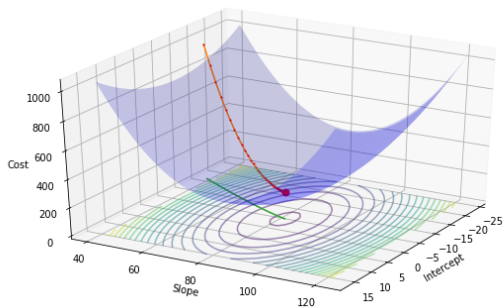
- Задача: найти минимум функции $J(\vec{w})$
- Инициализируем \vec{w}_0
- Посчитаем градиент. $\nabla J(\vec{w}_0)$
- $w_1 = w_0 - \alpha \nabla J(\vec{w}_0)$, α - скорость обучения

Градиентный спуск

- Задача: найти минимум функции $J(\vec{w})$
- Инициализируем \vec{w}_0
- Посчитаем градиент. $\nabla J(\vec{w}_0)$
- $w_1 = w_0 - \alpha \nabla J(\vec{w}_0)$, α - скорость обучения
- while True:
 $\vec{w}_{i+1} = \vec{w}_i - \alpha \nabla J(\vec{w}_i)$
 if $|\vec{w}_{i+1} - \vec{w}_i| < \epsilon$
 break



Результат



План лекции

1 Перцептрон

История развития

Перцептрон: введение

Логика на перцептроне

XOR проблема

2 Обучение перцептрона

Многослойные сети

Градиентный спуск

3 Различные активационные функции

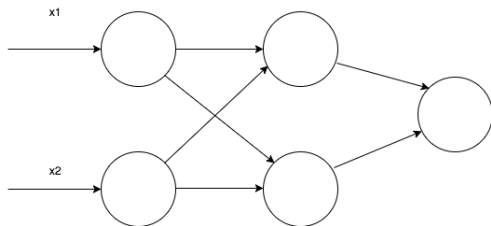
Друзья перцептрона

4 Обучение v.2

Правила обновления весов

Стохастический градиентный спуск

Линейный нейрон



Линейный нейрон

Активационная функция: $f(x) = x$



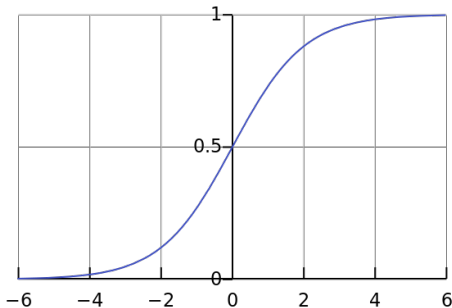
Сигмоидальный нейрон

Активационная
функция:

$$f(x) = \sigma(x)$$

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

Производная: $\sigma'(x) = ?$



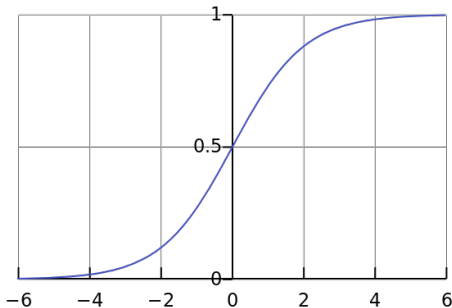
Сигмоидальный нейрон

Активационная
функция:

$$f(x) = \sigma(x)$$

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

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

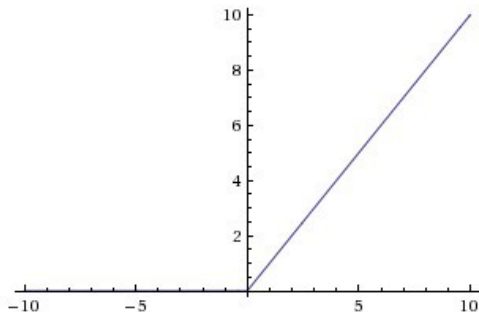


ReLU нейрон

Активационная
функция:

$$\text{relu}(x) = \max(0, x)$$

$$\text{relu}(x)' = ?$$



Гиперболический тангенс

Активационная функция:

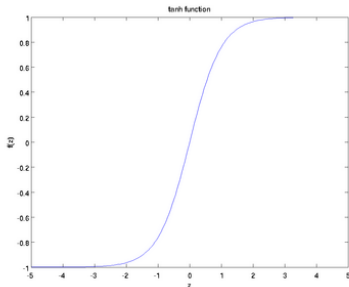
$$f(x) = th(x)$$

$$th(x) = \frac{\exp x - \exp -x}{\exp x + \exp -x}$$

Гиперболический тангенс.

Упражнение:

выразите $th(x)$ через $\sigma(x)$



План лекции

1 Перцептрон

История развития

Перцептрон: введение

Логика на перцептроне

XOR проблема

2 Обучение перцептрона

Многослойные сети

Градиентный спуск

3 Различные активационные функции

Друзья перцептрона

4 Обучение v.2

Правила обновления весов

Стохастический градиентный спуск

Градиенты целевых функций

Регрессия

$$\frac{\partial J}{\partial w_j} = \frac{1}{n} \sum_{i=1}^n (\hat{y}^{(i)} - y^{(i)}) x_j^{(i)}$$
$$\nabla J = \begin{pmatrix} \dots \\ \frac{\partial J}{\partial w_j} \\ \dots \end{pmatrix}$$

Классификация

$$\frac{\partial J}{\partial w_j} = \frac{1}{n} \sum_{i=1}^n (\sigma(w^T x^{(i)}) - y_i) \sigma(w^T x^{(i)}) (1 - \sigma(w^T x^{(i)})) x_j^{(i)}$$

План лекции

1 Перцептрон

История развития

Перцептрон: введение

Логика на перцептроне

XOR проблема

2 Обучение перцептрона

Многослойные сети

Градиентный спуск

3 Различные активационные функции

Друзья перцептрона

4 Обучение v.2

Правила обновления весов

Стохастический градиентный спуск

Стохастический градиентный спуск

- Задача: найти минимум функции $J(\vec{w})$
- Инициализируем точку w_0

Стохастический градиентный спуск

- Задача: найти минимум функции $J(\vec{w})$
- Инициализируем точку w_0
- Выберем случайно индекс t от 1 до n
- while True: Выберем случайно индекс t от 1 до n
 $w_{i+1} = w_i - \alpha \nabla J(w_i, x_t, y_t)$
 if $|w_{i+1} - w_i| < \epsilon$
 break

Mini-batch градиентный спуск

- Задача: найти минимум функции $J(\vec{w})$
- Инициализируем точку w_0
- w_0 Инициализация
-
- while True:
 Выберем случайно индекс $t_1 \cdots t_u$ от 1 до n

$$g_t = \sum_{c=1}^u \nabla J(w_i, x_c, y_c)$$

 $w_{i+1} = w_i - \alpha g_t$
 if $|w_{i+1} - w_i| < \epsilon$
 break

Градиентные методы: итог

Достоинства

- Допускает потоковое обучение(online)
- Позволяет найти экстремум не по всем данным
- Применяется к любым функциям потерь

Недостатки

- Локальная сходимость
- Подбор гиперпараметров
- Начальное приближение весов

IT'S SATURDAY NIGHT



**AND I'M SITTING HERE
LEARNING GRADIENT DESCENT**

- Посмотрели на перцептрон, как решение задачи классификации .
- Изучили простейшие методы оптимизации.
- Окунлись в целый мир активационных функций.
- Узнали как обучить простейший нейрон.

Литература I

На будущее можно почитать.



Y. Goodfellow.

Deeplearning Book.

MIT Press, 2017.



M. Nielsen

<https://neuralnetworksanddeeplearning.com/>



A. Moskvichev

<https://stepik.org/course/401>