

# Тятя! Тятя! Наши сети притащили мертвеца!

Ульянкин Филипп

## Аннотация

Однажды вечером Маша читала книгу.

Благодаря символическому дифференцированию вам никогда не придется заниматься реализацией алгоритма обратного распространения вручную. Поэтому не будем тратить время на его формулировку<sup>1</sup>.

Тут ещё пара таких цитат и рофл из Гудфелоу

**Доколе можно такое терпеть?** Если ты не понимаешь как самостоятельно реализовать самую топорную версию модели, а потом ввести в неё несколько простеньких улучшений — ты не понимаешь эту модель.

Написать по-нормальному abstract!

## Содержание

1	Всего лишь функция	2
2	50 оттенков градиентного спуска	6
3	Backpropagation	6
4	Активация	8
5	Регуляризаторы	9
6	Всего лишь кубики LEGO	9
7	Итоговый тест в стил Носко	9

Все должно быть в концепте маши кому принадлежит лёрнинг, все задачи должны ложиться в историю про неё, решения должны быть органично оторваны от личности Маши.

---

<sup>1</sup>Франсуа Шолле, Глубокое обучение на Python, стр. 77

# 1 Всего лишь функция

Ты всего лишь машина, только имитация жизни. Робот сочинит симфонию? Робот превратит кусок холста в шедевр искусства?

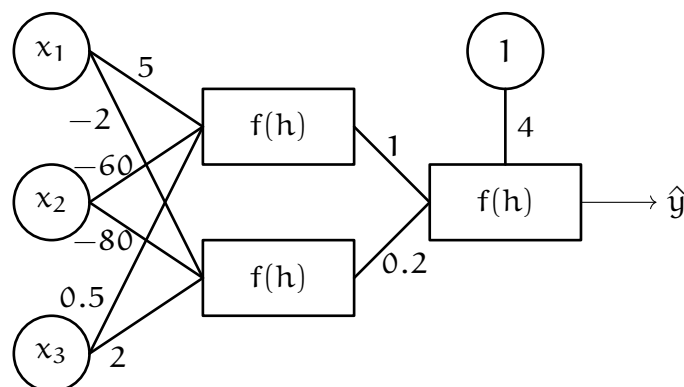
Из фильма «Я, робот» (2004)

## Упражнение 1 (от регрессии к нейросети)

Тут лёгкая задачка про то как регрессия превращается в нейронку + пункт про то что без функции активации будет то же самое на примере 2 регрессоров

## Упражнение 2 (из картинки в формулу)

Добродум хозяин кофейни на Тверской. Он хочет понять насколько сильно будет заполнена кофейня в следующие выходные. Для этого по старым данным он обучил нейросетку. На вход она принимает три фактора: температуру за окном,  $x_1$ , факт наличия на Тверской митинга,  $x_2$  и пол баристы на смене,  $x_3$ . В качестве функции активации Добродум использует ReLU.



- В эти выходные за барной стойкой стоит Агнесса. Митинга не предвидится, температура будет в районе 20 градусов. Сколько человек придёт в кофейню к Добродуму?
- На самом деле каждая нейросетка — это просто-напросто какая-то нелинейная сложная функция. Запишите нейросеть Добродума в виде функции.

## Упражнение 3 (из формулы в картинку)

Теперь в обратную сторону. Пусть у нас есть вот такая функция.

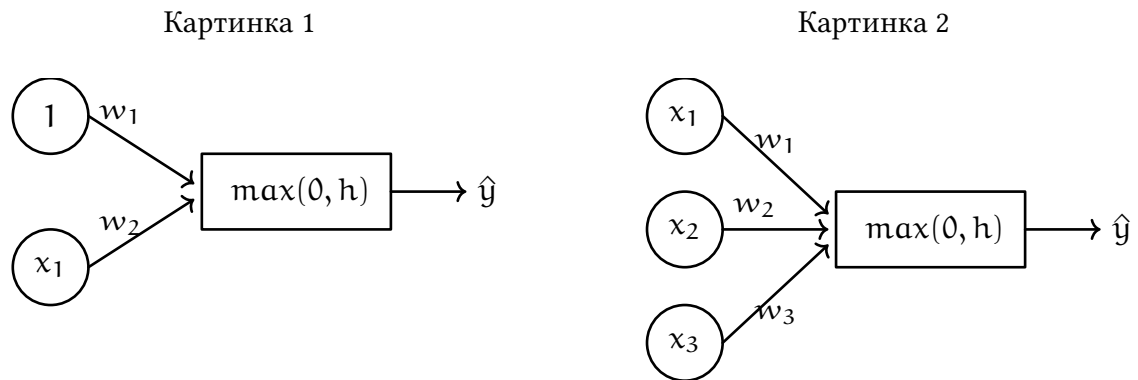
$$y = \max(0, 4 \cdot \max(0, 3 \cdot x_1 + 4 \cdot x_2 + 1) + 2 \cdot \max(0, 3 \cdot x_1 + 2 \cdot x_2 + 7) + 6)$$

Нарисуйте эту функцию в виде нейросетки.

## Упражнение 4 (армия регрессий)

Парни очень любят Машу, а Маша с недавних пор любит собирать персептроны и думать по вечерам об их весах и функциях активации. Сегодня она решила разобрать свои залежи из персептронов и как следует упорядочить их.

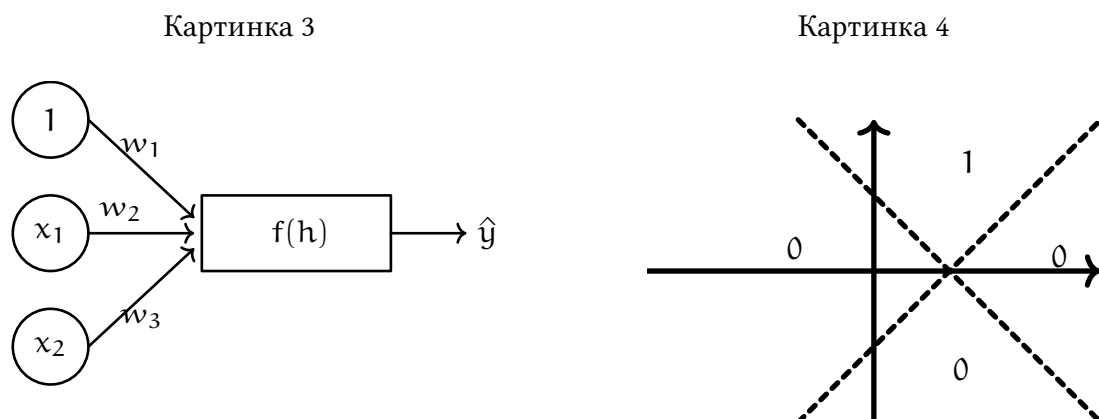
- а. В ящике стола Маша нашла персептрон с картинки 1 Маша хочет подобрать веса так, чтобы он реализовывал логическое отрицание, то есть превращал  $x_1 = 0$  в  $y = 1$ , а  $x_1 = 1$  в  $y = 0$ .



- б. В тумбочке среди носков Маша нашла персептрон, с картинки 2, Маша хочет подобрать такие веса  $w_i$ , чтобы персептрон превращал  $x$  из таблички в соответствующие  $y$ :

$x_1$	$x_2$	$x_3$	$y$
1	1	2	0.5
1	-1	1	0

- в. Оказывается, что в ванной всё это время валялась куча персептронов с картинки 3 с неизвестной функцией активации (надо самому выбрать).



Маша провела на плоскости две прямые:  $x_1 + x_2 = 1$  и  $x_1 - x_2 = 1$ . Она хочет собрать из персептронов нейросетку, которая будет классифицировать объекты с плоскости так, как показано на картинке 4.

- г. В коробке на кухне завалилось три персептрона, у каждого два входа с константой и пороговая функция активации. Маша хочет реализовать с их помощью функцию

$$y = \begin{cases} 1, & \text{если } x_2 \geq |x_1 - 3| + 2; \\ 0, & \text{иначе} \end{cases}.$$

## Упражнение 5 (логические функции)

Есть теорема, которая говорит о том, что с помощью нейросетки можно аппроксимировать почти любую функцию. Попробуйте с помощью нейросеток с минимально возможным числом нейронов описать логический функции, заданные следующими таблицами истинности:

$x_1$	$x_2$	$x_1 \cap x_2$
1	1	1
1	0	0
0	1	0
0	0	0

$x_1$	$x_2$	$x_1 \cup x_2$
1	1	1
1	0	1
0	1	1
0	0	0

$x_1$	$x_2$	$x_1 \text{ XoR } x_2$
1	1	0
1	0	1
0	1	1
0	0	0

Первые два столбика идут на вход, третий получается на выходе. Операция из третьей таблицы называется исключающим или (XoR). Если внимательно приглядеться, то можно заметить, что XoR — это то же самое что и  $[x_1 \neq x_2]^2$ .

## Упражнение 6 (ещё немного про XoR)

Маша заметила, что на XoR ушло очень много персептронов. Она поняла, что первые два персептрона пытаются сварить для третьего нелинейные признаки, которых нейросетке не хватает. Она решила самостоятельно добавить персептрону вход  $x_3 = x_1 \cdot x_2$  и реализовать XoR одним персептроном. Можно ли это сделать?

## Упражнение 7 (универсальный классификатор)

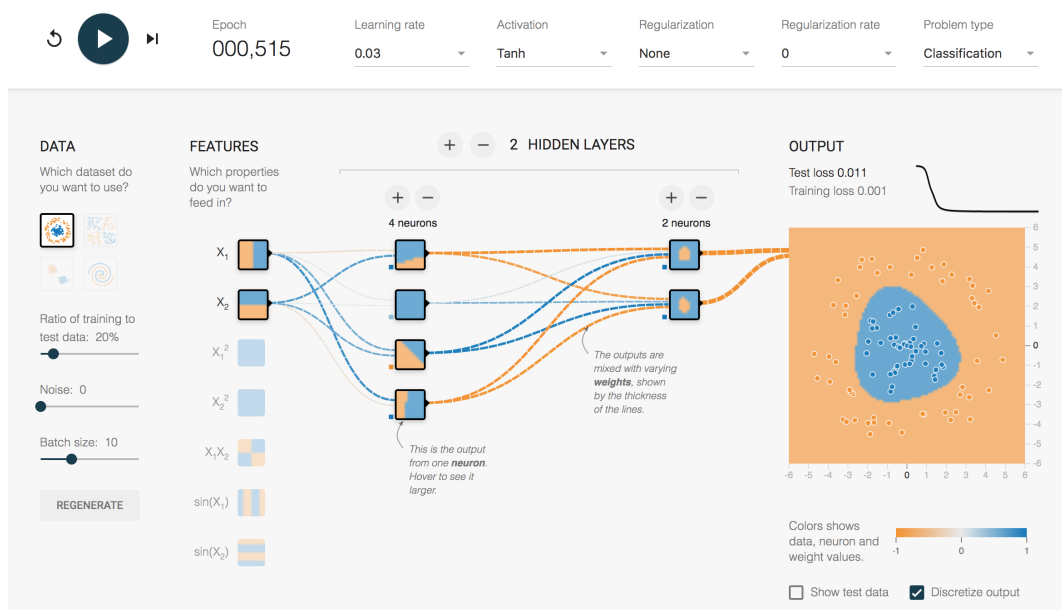
Маша задумалась о том, можно ли с помощью нейронной сетки с одним скрытым слоем и ступенчатой функцией активации решить абсолютно любую задачу классификации на два класса со сколь угодно большой точностью? Ей кажется, что да. Как это можно сделать?

## Упражнение 8 (избыток)

На сайте <http://playground.tensorflow.org> Маша стала играть с простенькими нейросетками и обучила для решения задачи классификации трёхслойного монстра.

---

<sup>2</sup>Тут квадратные скобки обозначают индикатор. Он выдаёт 1, если внутри него стоит правда и 0, если ложь.

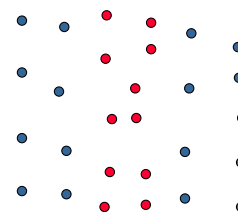
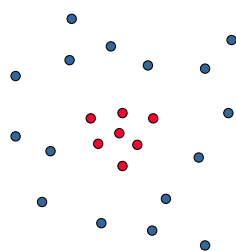


Голубым цветом обозначен первый класс, рыжим второй. Внутри каждого нейрона визуализирована та разделяющая поверхность, которую он выстраивает. Так, первый слой ищет разделяющую линию. Второй слой пытается из этих линий выстроить более сложные фигуры и так далее. Чем ярче связь между нейронами, тем больше весовой коэффициент, относящийся к ней. Синие связи — положительные, рыжие — отрицательные. Чем тусклее связь, тем он ближе к нулю.

Маша заметила, что с её получившейся архитектурой что-то не так. Какие проблемы вы в ней видите?

## Упражнение 9 (минималочка)

Сколько минимально нейронов необходимо для решения следующих задач классификации? Сколько слоёв минимально должно быть в нейросетке? Почему?



Добавить норм ксюр для точек как на семе

## Упражнение 10 (универсальный регрессор)

Доказать, что с помощью однослойной нейронной сетки можно приблизить любую непрерывную функцию от одного аргумента  $f(x)$  со сколь угодно большой точностью<sup>3</sup>.

**Hint:** Вспомните, что любую непрерывную функцию можно приблизить с помощью кусочно-линейной функции (ступеньки). Осознайте как с помощью пары нейронов можно описать такую

<sup>3</sup><http://neuralnetworksanddeeplearning.com/chap4.html>

ступеньку. Соедините все ступеньки в сумму с помощью выходного нейрона.

## 2 50 оттенков градиентного спуска

## 3 Backpropagation

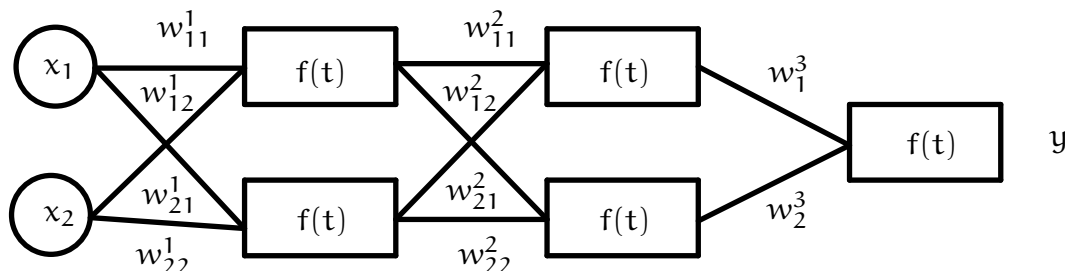
Что происходит, когда мы суём пальцы в розетку? Нас бьёт током! Мы делаем ошибку, и она распространяется по нашему телу назад.

### Упражнение 1 (граф вычислений)

Изобразите для функции  $f(x, y) = x^2 + xy + (x + y)^2$  граф вычислений. Найдите производные всех выходов по всем входам. Опираясь на граф выпишите частные производные функции  $f$ .<sup>4</sup>

### Упражнение 2 (придумываем backpropagation)

Дана нейросетка:

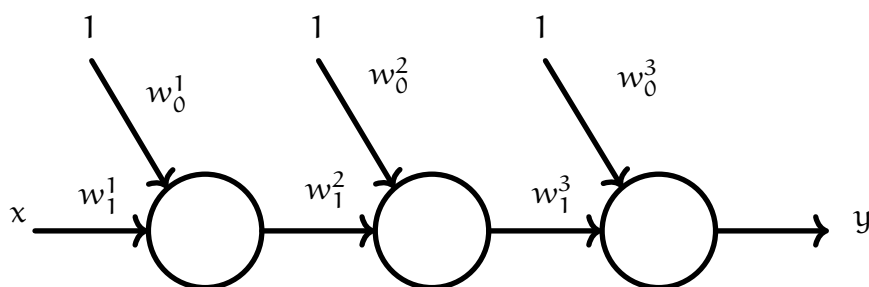


- Перепишите её как сложную функцию.
- Запишите эту функцию в матричном виде.
- Предположим, что  $L(W_1, W_2, W_3) = \frac{1}{2} \cdot (y - \hat{y})^2$  — функция потерь, где  $W_i$  — веса  $i$ -го слоя. Найдите производную функции  $L$  по всем весам  $W_i$ .
- Выглядит не очень оптимально, правда? Выпишите все производные в том виде, в котором их было бы удобно использовать для алгоритма обратного распространения ошибки, а затем, сформулируйте сам алгоритм.

### Упражнение 3 (backpropagation руками)

Как-то раз Вовочка решал задачу классификации. С тех пор у него в кармане завалялась нейросеть:

<sup>4</sup>По мотивам книги Николенко "Глубокое обучение"(стр. 79)



В качестве функции активации используется сигмоид:  $f(t) = \frac{e^t}{1+e^t}$ . Есть два наблюдения:  $x_1 = 1, x_2 = 5, y_1 = 1, y_2 = 0$ . Скорость обучения  $\gamma = 1$ . В качестве инициализации взяты нулевые веса. Как это обычно бывает, Вовочка обнаружил её в своих штанах после стирки и очень обрадовался. Теперь он собирается сделать два шага стохастического градиентного спуска, используя алгоритм обратного распространения ошибки. Помогите ему.

#### Упражнение 4 (ещё один backpropagation)

Пусть у нас есть нейронка:

$$y = f(X \cdot W_2) \cdot W_1$$

Как для функции потерь  $L(W_1, W_2) = (y - \hat{y})^2$  будет выглядеть алгоритм обратного распространения ошибки, если  $f(t) = \text{ReLU}(t) = \max(0; t)$ ? Найдите все выходы, все промежуточные производные. Опишите правило, по которому производная будет накапливаться, а также сам шаг градиентного спуска.

#### Упражнение 5 М

аша (ОПЯТЬ ОНА?!) собрала нейросеть:

$$y = \max \left( 0; X \cdot \begin{pmatrix} 1 & -1 \\ 0.5 & 0 \end{pmatrix} \right) \cdot \begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$$

Теперь Маша внимательно смотрит на неё.

- Первый слой нашей нейросетки — линейный. По какой формуле делается forward pass? Предположим, что на вход пришло наблюдение  $x = (1, 2)$ . Сделайте через этот слой forward pass и найдите выход из слоя.
- Найдите для первого слоя производную выхода по входу. При обратном движении по нейросетке, в первый слой пришёл накопленный градиент  $(-1, 0)$ . Каким будет новое накопленное значение градиента, которое выплюнет из себя линейный слой?
- Второй слой нейросетки — функция активации, ReLU. По какой формуле делается forward

pass? На вход в него поступило значение  $(2, -1)$ . Сделайте через него forward pass.

- г. Найдите для второго слоя производную выхода по входу. При обратном движении по нейросетке во второй слой пришёл накопленный градиент  $(-1, -2)$ . Каким будет новое накопленное значение градиента, которое выплюнет из себя ReLU?
- д. Третий слой нейросетки — линейный. По какой формуле делается forward pass? Пусть на вход поступило значение  $(2, 0)$ . Сделайте через него forward pass.
- е. Найдите для третьего слоя производную выхода по входу. При обратном движении по нейросетке, в третий слой пришёл накопленный градиент  $-2$ . Каким будет новое накопленное значение градиента, которое выплюнет из себя линейный слой?
- ж. Мы решаем задачу Регрессии. В качестве функции ошибки мы используем MSE. Пусть для рассматриваемого наблюдения реальное значение  $y = 0$ . Найдите значение MSE. Чему равна производная MSE по входу (прогнозу)? Каким будет накопленное значение градиента, которое MSE выплюнет из себя в предыдущий слой нейросетки, если изначально значение градиента инициализированно единицей?
- з. Пусть скорость обучения  $\gamma = 1$ . Сделайте для весов нейросети шаг градиентного спуска.

Посидела Маша, посидела, и поняла, что неправильно она всё делает. В реальности перед ней не задача регрессии, а задача классификации.

- а. Маша навинтила поверх второго линейного слоя сигмоиду. Как будет для неё выглядеть forward pass? Сделайте его. Найдите для сигмоиды производную выхода по входу.
- б. В качестве функции потерь Маша использует logloss. Как для этой функции потерь выглядит forward pass? Сделайте его. Найдите для logloss производную выхода по входу.
- в. Как будет выглядеть backward pass через logloss и сигмоиду? Прделайте его. Как изменится процедура градиентного спуска для остальной части сети?

## 4 Активация

Сюда перекинуть всё из task2

### Упражнение 1

Та, в чьих руках находится лёрнинг (это Маша), решила немного поэкспериментировать с выходами из своей сетки.

- а) Для начала Маша решила, что хочет решать задачу классификации на два класса и получать на выходе вероятность принадлежности к первому. Что ей надо сделать с последним слоем сетки?
- б) Теперь Маша хочет решать задачу классификации на  $K$  классов. Что ей делать с последним слоем?



- с) Новые вводные! Маша хочет спрогнозировать рейтинг фильма на "Кинопоиске". Он измеряется по шкале от 0 до 10 и принимает любое непрерывное значение. Как Маша может приспособить для этого свою нейронку?
- d) У Маши есть куча новостей. Каждая новость может быть спортивной, политической или экономической. Иногда новость может относиться сразу к нескольким категориям. Как Маше собрать нейросетку для решения этой задачи? Как будет выглядеть при этом функция ошибки?
- e) Маша пошла в кафе. А там куча народу. Сейчас она сидит за столиком, попивает ванильный топлёный кортадо и думает о нём, о лёрнинге. Сейчас мысль такая: как можно спрогнозировать число людей в кафе так, чтобы на выходе сетка всегда прогнозировала целое число. Надо ли как-то при этом менять функцию потерь?

## 5 Регуляризаторы

## 6 Всего лишь кубики LEGO

## 7 Итоговый тест в стел Носко

- a. Вопрос про батчнормализацию первым слоем вместо нормализации в предобработке.