

Тятя! Тятя! Наши сети притащили мертвеца!

Ульянкин Филипп

Аннотация

В этой виньетке собрана коллекция ручных задачек про нейросетки на пару томных вечеров. Вместе с Машей можно попробовать по маленьким шажкам с ручкой и бумажкой раскрыть у себя в теле несколько чакр и немного глубже понять модели глубокого обучения.

Вместо введения

Однажды Маша услышала про какой-то Машин лёрнинг. Она сразу же смекнула, что именно она та самая Маша, кому этот лёрнинг должен принадлежать. Ещё она смекнула, что если хочет владеть лёрнингом по праву, ни одна живая душа не должна сомневаться в том, что она шарит. Поэтому она постоянно изучает что-то новое.

Её друг Миша захотел стать адептом Машиного лёрнинга, и спросил её о том, как можно за вечер зашарить алгоритм обратного распространения ошибки. Тогда Маша открыла свою первую книгу по глубокому обучению и прочитала в ней:

Благодаря символическому дифференцированию вам никогда не придется заниматься реализацией агоритма обратного распространения вручную. Поэтому не будем тратить время на его формулировку¹.

Маше такая логика показалась странной. Поэтому она взяла книгу с более глубокой математикой. Там она прочитала, что:

Николенко

Тогда Маша взяла Библию глубокого обучения² и поняла, что по ней за один вечер точно не разберёшься. Слишком серьёзно всё написано. Для вечерних разборок нужно что-то более инфантильное.

У Маши оставался один выход: поскрести по лёрнингу и собрать инфантильную коллекцию ручных задачек, прорешивая которую новые адепты Машиного лёрнинга могли бы открывать у себя во чакру за чакрой. Так и появилась эта виньетка.

¹Франсуа Шолле, Глубокое обучение на Python, стр. 77

²Goodfellow I., Bengio Y., Courville A. Deep learning. – MIT press, 2016.

Содержание

1	Всего лишь функция	3
2	50 оттенков градиентного спуска	8
3	Backpropagation	8
4	Активация	10
5	Регуляризаторы	12
6	Всего лишь кубики LEGO	12
7	Итоговый тест в стел Носко	12

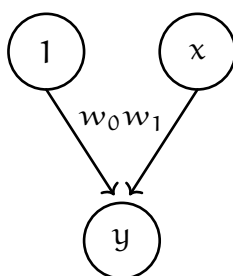
1 Всего лишь функция

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

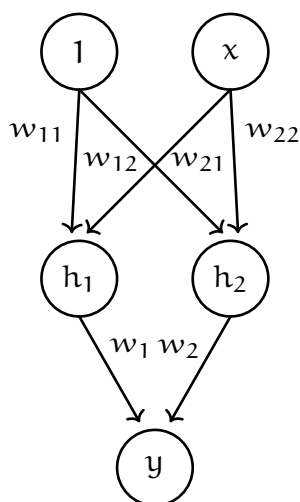
Из фильма «Я, робот» (2004)

Упражнение 1 (от регрессии к нейросетке)

Однажды вечером, по пути с работы³ Маша зашла в свою любимую кофейню на Тверской. Там, на стене, она обнаружила очень интересную картину:



Хозяин кофейни, Добродум, объяснил Маше, что это Покрас-Лампас так нарисовал линейную регрессию,⁴ и её легко можно переписать в виде формулы: $y_i = w_0 + w_1 \cdot x_i$. Пока Добродум готовил кофе, Маша накидала у себя на бумажке новую картинку:



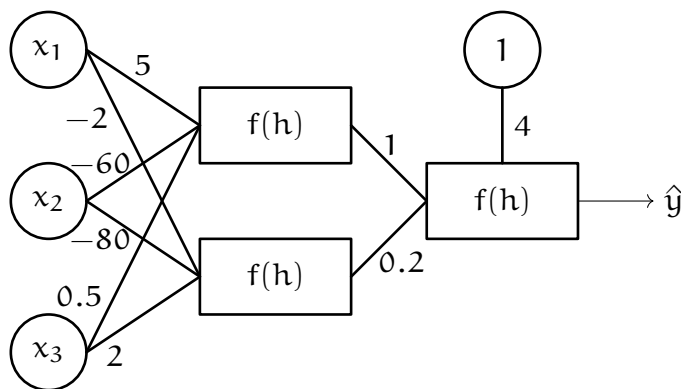
- Как такая функция будет выглядеть в виде формулы?
- Правда ли, что y будет нелинейно зависеть от x ?
- Если нет, как это исправить и сделать зависимость нелинейной?

³она работает рисёрчером.

⁴эксклюзивный заказ был

Упражнение 2 (из картинки в формулу)

Добродум хочет понять насколько сильно будет заполнена кофейня в следующие выходные. Для этого 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 барной⁵ стойкой стоит Агнесса. Митинга не предвидится, температура будет в районе 20 градусов. Сколько человек придёт в кофейню к Добродуму?
- На самом деле каждая нейросеть — это просто-напросто какая-то нелинейная сложная функция. Запишите нейросеть Добродума в виде функции.

Упражнение 3 (из формулы в картинку)

Маша написала на бумажке функцию:

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

Теперь она хочет, чтобы кто-нибудь из её адептов нарисовал её в виде нейросетки. Нарисуй.

Упражнение 4 (армия регрессий)

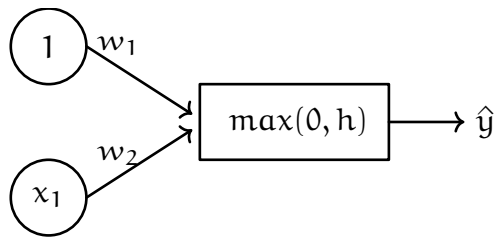
Парни очень любят Машу,⁶ а Маша с недавних пор любит собирать персептроны и думать по вечерам об их весах и функциях активации. Сегодня она решила разобрать свои залежи из персептронов и как следует упорядочить их.

- 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 $y = 0$ (так работает логическая функция: отрицание).

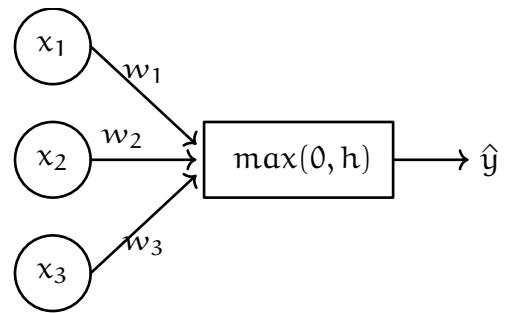
⁵барной... конечно, кофейня у него...

⁶когда у тебя есть лёрнинг, они так и лезут

Картинка 1



Картинка 2

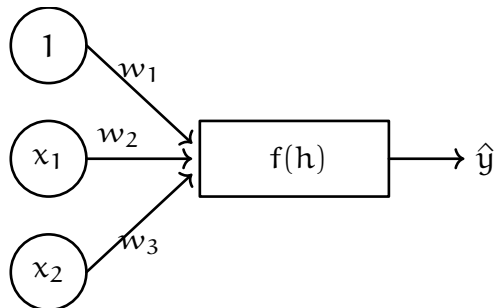


- б. В тумбочке, среди носков, Маша нашла персептрон, с картинки 2, Маша хочет подобрать такие веса w_i , чтобы персептрон превращал x из таблички в соответствующие y :

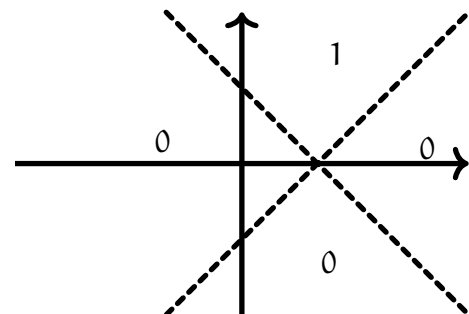
x_1	x_2	x_3	y
1	1	2	0.5
1	-1	1	0

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

Картинка 3



Картинка 4



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.

Упражнение 5 (логические функции)

Маша вчера поссорилась с Пашей. Он сказал, что у неё нет логики. Чтобы доказать Паше обратное, Маша нашла теорему, которая говорит о том, что с помощью нейросетки можно аппроксимировать почти любую функцию, и теперь собирается заняться аппроксимацией логических функций. Для начала она взяла самые простые, 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. Первая операция — логическое "и" вторая — "или". Операция из третьей таблицы называется "исключающим или" (XoR). Если внимательно приглядеться, то можно заметить, что XoR — это то же самое что и $[x_1 \neq x_2]$ ⁷.

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)

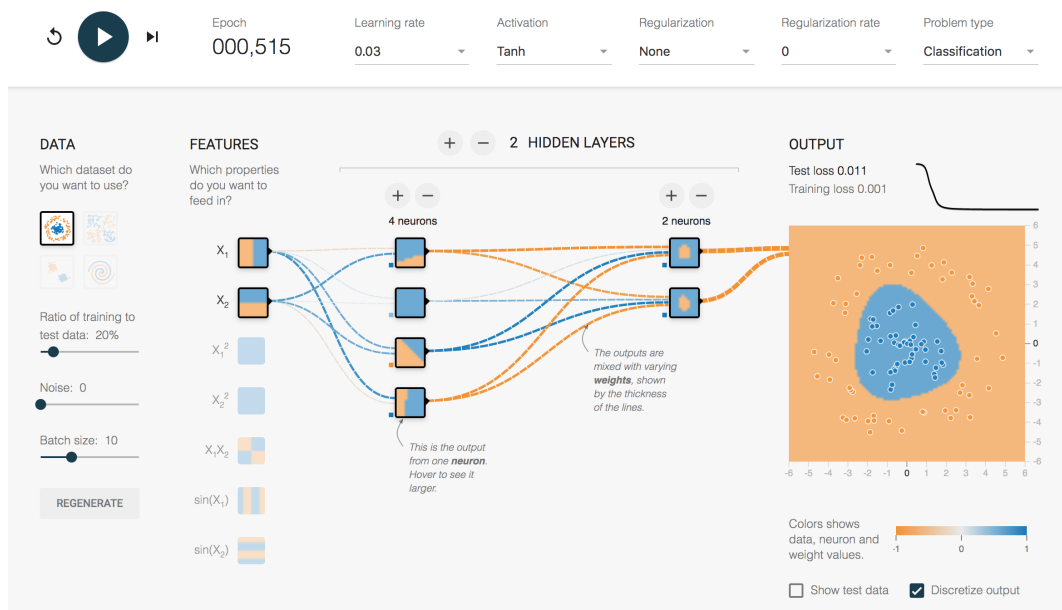
Маша задумалась о том, можно ли с помощью нейронной сетки с одним скрытым слоем и ступенчатой функцией активации решить абсолютно любую задачу классификации на два класса со сколь угодно большой точностью. Ей кажется, что да. Как это можно сделать?

Упражнение 8 (избыток)

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

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.

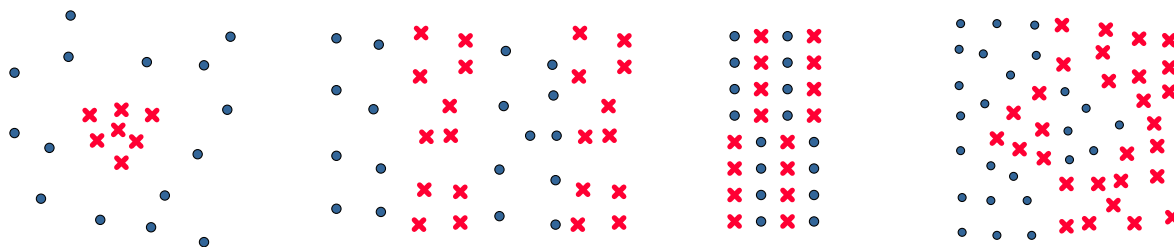
⁷Тут квадратные скобки обозначают индикатор. Он выдаёт 1, если внутри него стоит правда и 0, если ложь.



Маша заметила, что с её архитектурой что-то не так. Какие у неё проблемы?

Упражнение 9 (минималочка)

Шестилетняя сестрёнка ворвалась в квартиру Маши и разрисовала ей все обои:



Маша по жизни оптимистка. Поэтому она увидела не дополнительные траты на ремонт, а четыре задачи классификации. И теперь в её голове вопрос: сколько минимально нейронов нужно, чтобы эти задачи решить?

Упражнение 10 (универсальный регрессор)

Маша доказала Паше, что у неё всё в полном порядке с логикой. Теперь она собирается доказать ему, что с помощью однослойной нейронной сети можно приблизить любую непрерывную функцию от одного аргумента $f(x)$ со сколь угодно большой точностью⁸.

Hint: Вспомните, что любую непрерывную функцию можно приблизить с помощью кусочно-линейной функции (ступеньки). Осознайте как с помощью пары нейронов можно описать такую ступеньку. Соедините все ступеньки в сумму с помощью выходного нейрона.

⁸<http://neuralnetworksanddeeplearning.com/chap4.html>

2 50 оттенков градиентного спуска

3 Backpropagation

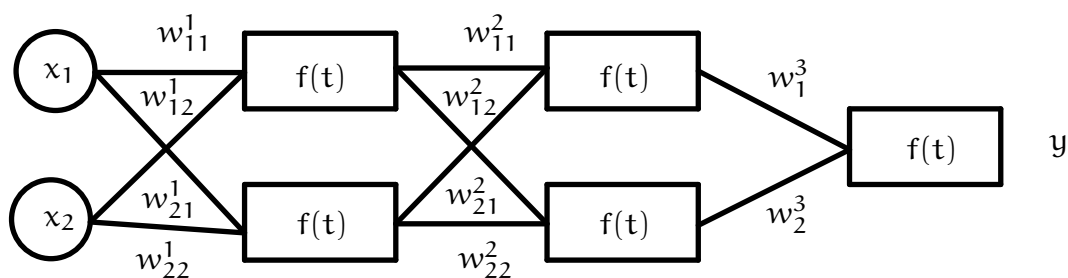
Что происходит, когда мы суём пальцы в розетку? Нас бьёт током! Мы делаем ошибку, и она распространяется по нашему телу назад.

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 .⁹

Упражнение 2 (придумываем backpropagation)

Дана нейросетка:

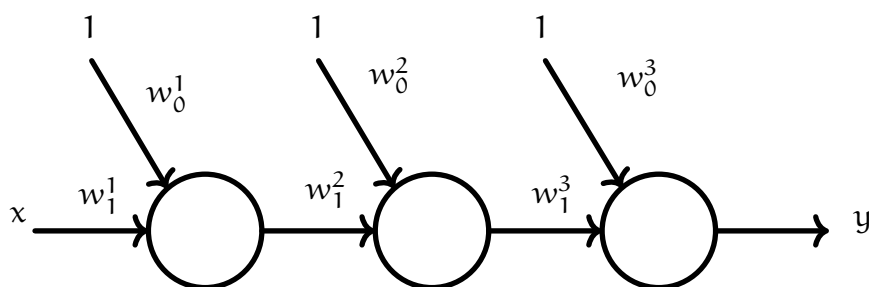


- Перепишите её как сложную функцию.
- Запишите эту функцию в матричном виде.
- Предположим, что $L(W_1, W_2, W_3) = \frac{1}{2} \cdot (y - \hat{y})^2$ — функция потерь, где W_i — веса i -го слоя. Найдите производную функции L по всем весам W_i .
- Выглядит не очень оптимально, правда? Выпишите все производные 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:

⁹По 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 выплунет 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 выплунет 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 выплунет 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

Упражнение 1 У

Бандерлога три наблюдения¹⁰, первое наблюдение — кит, остальные — муравьи. Киты кодируются $y_i = 1$, муравьи — $y_i = 0$. В качестве регрессоров Бандерлог берёт номера наблюдений $x_i = i$. После этого Бандерлог оценивает логистическую регрессию с константой.

- а. Выпишите эмпирическую функцию риска, которую минимизирует Бандерлог;
- б. При каких оценках коэффициентов логистической регрессии эта функция достигает своего минимума?

Упражнение 2

¹⁰Про другие приключения Бандерлога читай тут: https://github.com/bdemeshev/mlearn_pro/blob/master/mlearn_pro.pdf

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

- a) 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?
- b) Теперь Маша хочет решать задачу классификации на K классов. Что ей делать с последним слоем?
- c) 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 нейросеть 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?

Упражнение 3 Б

андерлог чуть внимательнее присмотрелся к своему третьему наблюдению и понял, что это не кит, а бобёр. Теперь ему нужно решать задачу классификации на три класса. Он решил использовать для этого нейросеть с softmax-слоем на выходе. Предположим, что сетка обучилась и на двух новых наблюдениях, перед самым softmax-слоем она выплюнула 1, 2, 5 и 2, 5, 1.

- a. Чему равны вероятности получить кита, муравья и бобра для обеих ситуаций?
- б. Пусть первым был кит, а вторым бобёр. Чему будет равна logloss-ошибка?

Упражнение 4 И

Когда в функцию Softmax добавляют дополнительный параметр T , который называют температурой. Тогда она приобретает вид

$$f(z) = \frac{e^{\frac{z_i}{T}}}{\sum_{k=1}^K e^{\frac{z_k}{T}}}$$

Обычно это делается, когда с помощью нейросетки нужно сгенерировать какой-нибудь новый объект. Пусть у нас есть три класса. Наша нейросеть выдала на последнем слое числа 1, 2, 5.

- a. Какое итоговое распределение вероятностей мы получим, если $T = 10$?
- б. А если $T = 1$?

- в. А если $T = 0.1$?
- г. Какое распределение получится при $T \rightarrow 0$?
- д. А при $T \rightarrow \infty$?
- е. Предположим, что объектов на порядок больше. Например, это реплики, которые Алиса может сказать вам в ответ на какую-то фразу. Понятное дело, что вашей фразе будет релевантно какое-то подмножество ответов. Какое значение температуры сэмплирования T смогут сделать реплики Алисы непредсказуемыми? А какие сделают их однотипными?

Упражнение 5 Ф

функция $f(t) = \frac{e^t}{1+e^t}$ называется сигмой¹¹.

- а. Что происходит при $t \rightarrow +\infty$? А при $t \rightarrow -\infty$?
- б. Как связаны между собой $f(t)$ и $f(-t)$?
- в. Как связаны между собой $f'(t)$ и $f'(-t)$?
- г. Как связаны между собой $f(t)$ и $f'(t)$?
- д. Найдите $f(0)$, $f'(0)$ и $\ln f(0)$.
- е. Найдите обратную функцию $f^{-1}(t)$
- ж. Как связаны между собой $\frac{d \ln f(t)}{dt}$ и $f(-t)$?
- з. Постройте графики функций $f(t)$ и $f'(t)$.
- и. Разложите $h(\beta_1, \beta_2) = \ln f(y_i(\beta_1 + \beta_2 x_i))$ в ряд Тейлора до второго порядка в окрестности точки $\beta_1 = 0, \beta_2 = 0$.

5 Регуляризаторы

6 Всего лишь кубики LEGO

7 Итоговый тест в стел Носко

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

¹¹В этом всё тоже замешан один Бандерлог.