

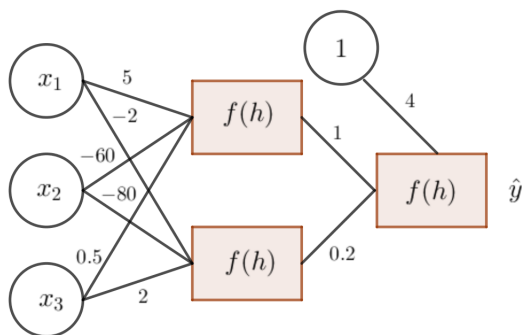
Тятя! Тятя! Наши сети притащили мертвеца!

эконом РАНХиГС
осень 2019

Задачи к посиделке 1

Задача 1

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 , пол баристы на смене, x_2 и факт наличия на Тверской митинга, x_3 . В качестве функции активации Добродум использует ReLU.



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

Задача 2

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

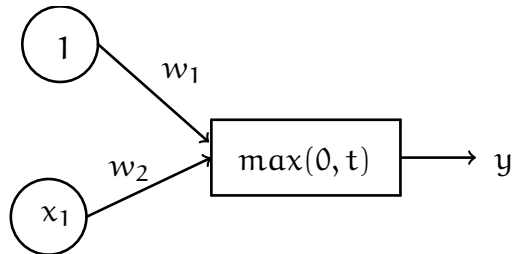
$$y = \max(0, 4 \cdot \max(0, 3x_1 + 4x_2 + 1) + 2 \cdot \max(0, 3x_1 + 2x_2 + 7) + 6)$$

Нарисуйте эту функцию в виде нейросетки.

Задача 3

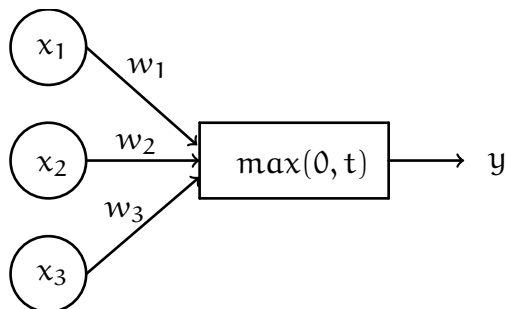
Парни очень любят Свету, а Света любит собирать перцептроны и думать по вечерам об их весах и функциях активации. Сегодня она решила разобрать свои залежи из перцептронов и как следует упорядочить их.

- Для перцептрона



нужно подобрать веса так, чтобы он превращал $x_1 = 0$ в $y = 1$, а $x_1 = 1$ в $y = 0$.

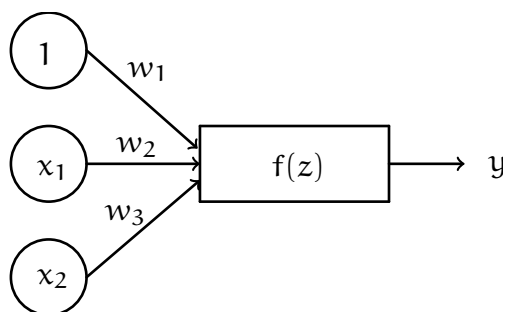
- Для перцепторона



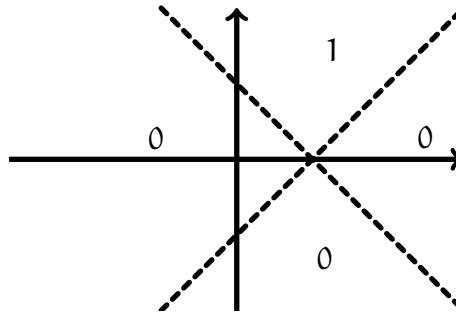
Света хочет по наблюдениям x подобрать такие веса w_i , чтобы на выходе получились y .

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

- У Светы есть несколько вот таких перцептронов с неизвестной функцией активации (надо самому выбирать):



На плоскости проведены две прямые $x_1 + x_2 = 1$ и $x_1 - x_2 = 1$.



Свете нужно собрать нейросетку, которая будет классифицировать объекты с плоскости так, как показано на картинке.

Задача 4

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

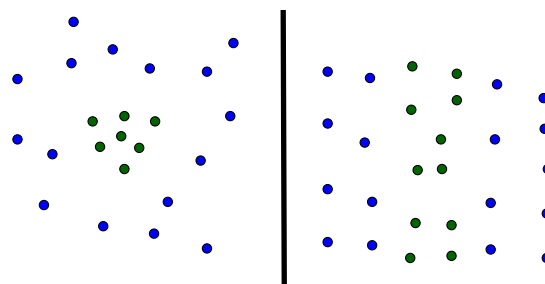
x_1	x_2	$x_1 \cup x_2$
1	1	1
1	0	1
0	1	1
0	0	0

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.

Задача 5

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



Задача 6

Доказать, что с помощью однослойной нейронной сетки можно приблизить любую непрерывную функцию от одного аргумента $f(x)$ со сколь угодно большой точностью¹.

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.

Задача 7

Маша услышала про машин лёрнинг и решила, что они и есть та самая Маша, которой этот лёрнинг принадлежит. Теперь она собирается обучить нейронную сеть для решения задачи регрессии. На вход в неё идёт 12 переменных, в сетке есть 3 скрытых слоя. В первом слое 300 нейронов, во втором 200, в третьем 100.

- a) Сколько параметров предстоит оценить Маше? Сколько наблюдений вы бы на её месте использовали?
- b) Пусть в каждом слое была отключена половина нейронов. Сколько коэффициентов необходимо оценить?
- c) Предположим, что Маша решила после первого слоя добавить в свою сетку Dropout с вероятностью p . Какова вероятность того, что отключится весь слой?
- d) Маша добавила Dropout с вероятностью p после каждого слоя. Какова вероятность того, что один из слоёв отключится и сетка не сможет учиться?
- e) Пусть случайная величина N — это число включённых нейронов. Найдите её математическое ожидание и дисперсию. Если Маша хочет проредить сетку на четверть, какое значение p она должна поставить?
- f) Пусть случайная величина P — это число параметров в нейросети, которое необходимо оценить. Найдите её математическое ожидание и дисперсию. Почему найденное вами математическое ожидание выглядит очень логично? Что оно вам напоминает? Обратите внимание на то, что смерть одного из параметров легко может привести к смерти другого.

Задача 8

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?

¹<http://neuralnetworksanddeeplearning.com/chap4.html>

- д) 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?
- е) 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?

Ещё задача

Если вы никогда не решали эту задачку, рекомендую сделать это. Она довольно неплохо открывает чакру на работу с регуляризаторами, и показывает как именно они стягивают к нулю коэффициенты, не давая модели переобучаться. В рамках нейросеток механизм точно такой же.

Задача 9

Вася измерил вес трёх покемонов, $y_1 = 6$, $y_2 = 6$, $y_3 = 10$. Вася хочет спрогнозировать вес следующего покемона. Модель для веса покемонов у Васи очень простая, $y_i = \beta + \varepsilon_i$, поэтому прогнозирует Вася по формуле $\hat{y}_i = \hat{\beta}$.

Для оценки параметра β Вася использует следующую целевую функцию:

$$\sum (y_i - \hat{\beta})^2 + \lambda \cdot \hat{\beta}^2$$

- а) Найдите оптимальное $\hat{\beta}$ при $\lambda = 0$.
- б) Найдите оптимальное $\hat{\beta}$ при произвольном λ . Правда ли, что чем больше λ , тем меньше β ?
- в) Подберите оптимальное λ с помощью кросс-валидации *leave one out* («выкинь одного»). При такой валидации на первом шаге мы оцениваем модель на всей выборке без первого наблюдения, а на первом тестируем её. На втором шаге мы оцениваем модель на всей выборке без второго наблюдения, а на втором тестируем её. И так далее n раз. Каждое наблюдение является отдельным фолдом.
- г) Найдите оптимальное $\hat{\beta}$ при λ_{CV} .