

# Нейронки. Лекция 1. Лисичка

Геруня солнышко

09.04.2024

## 1 Интродукшн

Нейронки, как и все хорошее люди спиздили у природы. сейчас мы построим математический объект, который пиздат тем, что может использоваться для решения задач нерешаемых детерминированными алгоритмами.

## 2 Вводные по нейронным сетям

Нервная система человека состоит из нейронов и нервных волокон, передающих сигналы. И весь твой пердеж и пиздеж непосредственно зависит от того, как бегают сигналы по тельцу. Нейронная сеть - математическая модель построенная на математическом взаимодействии нейронов и функционирующая по принципам процессов протекающих в мозге человека. Характерной особенностью является многослойная структура, в которой есть два особых (нетрадиционной сексуальной ориентации) слоя входной и выходной.

Функционирование нейронной сети - это процесс обработки сигналов (символов), поступающих на нейроны входного слоя и вычисление выходных сигналов нейронов выходного слоя.

Типикал задачи:

- классификация (кто тупой, а кто совсем дибил)
- прогнозирование
- распознавание образа

Дефолтные архитектуры:

- перцептрон (от англ. perception - восприятие)
- сверточная нейронная сеть

Перцептрон  $:= G(X, U)$  - оргграф на множестве вершин  $X$  со множеством стрелок  $U$

Тут  $v \in X$  называют нейронами.  $X_{in} \subset X$  - множество нейронов входного слоя,

$X_{out} \subset X$  - множество нейронов выходного слоя,

а  $X \setminus (X_{in} \cup X_{out})$  - нейроны скрытых слоев.

Перцептрон - полносвязный граф  $\Rightarrow \forall x \in X_j \exists e(x, y) \in U \forall y \in (X_{j-1} \cup X_{j+1})$ , где  $X_j$  это множество вершины  $j$ -го слоя

Для каждого ребра определим вес:  $\forall u \in U : w(u) \in [-1; 1]$  - на ребро нулевого веса всем похуй, чем больше модуль веса, тем больше влияние этого ребра на сигнал (положительное или отрицательное)

Обозначим за  $input(x), output(x)$  сумму весов входящих и исходящих стрелок для вершины  $x$ .

Заметим, что  $\forall x \in X_{in} : input(x) = output(x)$  (сколько втекло, столько вытекло), а

$\forall x \in X_j \in (X \setminus X_{in}) : input(x) = \sum_{\forall x_i \in X_{j-1}} output(x_i) \cdot W_{ix}$  то есть входной сигнал нейрона, это

сумма выходных сигналов вершин стрелки, которых ведут в  $x$ , умноженных на веса соответствующих ребёр.

$output(x) = f(input(x))$  где  $f$  - функция активации нейрона  $x$ .

Типичные функции активации:

- сигмовидная

$$f(a) = \frac{1}{1 + e^{-a}}$$

- линейная

$$f(a) = a$$

- гиперболический тангенс

$$f(a) = \frac{e^{2a} - 1}{e^{2a} + 1}$$

- линейный выпрямитель ( $RELU$ )  $\begin{cases} f(a) = a, & \text{if } a > 0 \\ f(a) = 0, & \text{if } a \leq 0 \end{cases}$

Для оценивания полученного решения используются разные способы. Оценка, как правило, явно зависит от выходных сигналов нейронной сети и неявно (через функционирование) от всех других ее сигналов.

В качестве метрики базового варианта можно использовать MSE.

$$MSE := \frac{1}{n_{out}} \cdot \sum_{\forall x \in X_{out}} \left( output(x) - output^*(x) \right)^2$$

Тут  $output^*(x)$  - ожидаемые выходные сигналы

### 3 Обучение нейронной сети

Нейронные сети подобно человеческим мозгам обучаемы, но вопрос как?

Человеки способны к самообучению, и могут достичь успехов, не зная природы процессов, лежащих в основе их действий. Непременными атрибутами обучения двуногих являются многократное повторение и моментальная оценка действий.

Виды обучений:

- обучение с учителем
- обучение без учителя
- обучение с подкреплением
- обучение во ВШЭ (не рекомендуется)

Цель обучения с учителем, состоит в подборе таких весов для ребёр, чтобы при заданных значениях входных сигналов получить близкие к ожидаемым (правильным) выходные сигналы.

Набор входных сигналов и правильных ответов называют обучающим примером.

Обучающая выборка - совокупность нескольких обучающих примеров.

Для поиска весов стрелок в нейронной сети требуется прежде всего критерий, который будет

определять расстояние между правильными сигналами и полученными, то есть некоторая функция  $E(W)$ , где  $W := (w_1, w_2, \dots, w_{|U|})$  - множество весов стрелок сети

Замечание:  $E(W) \in C^1[-1;1]$  другими словами функция критерия непрерывна и дифференцируема на отрезке  $[-1;1]$

$\square E^1(W), E^2(W), \dots, E^k(W)$  - последовательность значений критерия на каждом шаге обучения, чтоб наше обучение не превратилось в университет Синергии было бы славно, чтоб такая последовательность была убывающей

Тогда пойдем, когда пора остановить обучение, возможности такие:

$k_{max}$  то есть заданное наперед количество шагов обучения, или

$$\left| E^k(W) - E^{k-1}(W) \right| \leq \varepsilon \text{ заданная наперед точность сети}$$

$$W_{ij}^k = W_{ij}^{k-1} + \Delta W_{ij}^k \quad (2)$$

формула для весов на  $k$ -ом шаге обучения, как видите она зависит от весов на предыдущем шаге и некоторого  $\Delta W_{ij}^k$ , а оно в свою очередь вычисляется так:

$$\Delta W_{ij}^k = h^k \cdot p_{ij}^k \quad (3)$$

тут  $p_{ij}^k$  вектор в многомерном пространстве, а  $h^k$  это шаг в этом направлении

$\frac{dE}{dW}$  - первая производная целевой функции, поэтому для стремления к минимуму надо выбрать направление антиградиента.

Для нейронов выходного слоя эта задача решается градиентными методами относительно легко

А для остальных слоев используется метод обратного распространения ошибки. О котором будет рассказано на следующей лекции, которую я вряд-ли посетую.

## 4 2ая лекция

хаха сосать я прогулял английский и всё-таки пришел на 2-ую лекцию.

### 4.1 Метод обратного распространения ошибки для обучения НС

Ошибки это обязательный момент обучения, а вот нахуя их распространять не чилл вопросик. Помните сигмовидную функцию, она есть выше заметим, что:

$$f'(a) = f(a) \cdot (1 - f(a)) \quad (4)$$

здесь  $a$  входной сигнал, напомним его формулу:

$$\forall x \in X_j \in (X \setminus X_{in}) : input(x) = \sum_{\forall x_i \in X_{j-1}} output(x_i) \cdot W_{ix} \quad (5)$$

входной сигнал нейрона, это сумма выходных сигналов вершин стрелки, которых ведут в  $x$ , умноженных на веса соответствующих ребёр.

Функцию оценки решения зададим во какую

$$E(w) = \frac{1}{2} \cdot \sum_{\forall x \in X_{out}} (output(x) - output^*(x))^2 \quad (6)$$

это очень похоже на MSE, формулу которого мы уже приводили, но вы не понимаете это другое

Ща глянем на производную этой ебанины за какаим-то хуем

$$\frac{dE}{dw} = (output(x) - output^*(x)) \cdot \frac{df}{da} \cdot \frac{da}{dw} \quad (7)$$

Так как высшая цель нашего существования это уменьшение ошибки, то нам надо въебать минусик, чтоб получить направление антиградиента

$$-\frac{dE}{dw} = (output^*(x) - output(x)) \cdot \frac{df}{da} \cdot \frac{da}{dw}$$

Подставим туда формулу (4) чтоб жизнь малиной не казалась

$$-\frac{dE}{dw} = (output^*(x) - output(x)) \cdot output(x) \cdot (1 - output(x)) \cdot \frac{da}{dw} \quad (8)$$

Для нейронов выходного слоя у которых нет исходящих стрелок (функция  $E$  не зависит от  $\frac{da}{dw}$ ) оценим ошибку, которая допущена на данном нейроне как:

$$\forall x \in X_{out} \delta(x) = (output^*(x) - output(x)) \cdot output(x) \cdot (1 - output(x)) \quad (9)$$

Возник вопрос(сколько у цыганки на пизде волос) как посчитать  $(output^*(x) - output(x))$  для нейронов НЕвыходного слоя, если мы не знаем для них правильные ответы( $output^*(x)$ ).

Ответ: можно оценить вот такой хуйтой:

$$\sum_{x \in Im(x)} \delta(x) \cdot W_{js}$$

где  $Im(X)$  это множество образов (а.к.а. множество вершин следующего слоя) а индекс  $W$  пробегает все стрелки из  $x$  в вершины след. слоя

$$\forall x \in (X \setminus (X_{out} \cup X_{in})) \delta(x) = \sum_{x \in Im(x)} \delta(x) \cdot W_{js} \cdot output(x) \cdot (1 - output(x)) \quad (10)$$

Таким образом вектора антиградиента в формуле (8) заменяются на  $\delta(x) \cdot \frac{da}{dw}$  (11)  
Дифференцием (5) и получим формулу для каждого нахуй компонента вектора градиента

$$\rho^k(W_{ij}) = \delta(x) \cdot output(x) \quad (12)$$

это есть вектор указывающий нам верный путь в светлое будущее на  $k$ -ом шаге обучения

$$\Delta W_{ij}^k = \mu \cdot \rho(W_{ij}) + \alpha \cdot \Delta W_{ij}^{k-1} \quad (13)$$

если вы забыли, что за  $\Delta W_{ij}^k$ , то гляньте формулу (3)

$\mu$  это скорость - аналог  $h$  из формулы (3) а  $\alpha$  это момент обучения - специальный коэффициент задаваемый учителем, чтоб подогладить нашу функцию

## 5 Свёрточные нейронные сети а.к.а СНС

Такую архитектуру предложил некто Я. Лекун (а я лизун) в 1988г. для решения задач на распознавание образов.

Идея основана на особенностях зрительной коры головного мозга человека (как обычно спиздили) в которой есть 2 типа нейронов - простые и сложные. Простые реагируют на прямые линии под разным углом, а затем сложные активируют некоторые наборы простых. Поэтому структура СНС - многослойный ориентированный граф, в котором между входным и выходным слоем чередуются сверточные слои и слои выборки.

Входной слой есть матрица  $n_1 \times n_1$  - изображение закодированное циферками

Функционирование СНС - переход от конкретных особенностей изображения к все более и более абстрактным деталям, при этом сеть самонастраивается и сама вырабатывает иерархию абстрактных признаков (последовательных карт признаков). В отличие от перцептрона (где у каждой стрелки свой вес) вычисление выходного сигнала нейрона происходит с помощью операции свертки, для которой разработчик задает ядра свертки -  $Y_1, Y_2, \dots, Y_k$  набор небольших квадратных матриц, где  $k \leq n_2$ , а  $n_2$  сильно меньше  $n_1$

Пусть есть исходная матрица  $M$  размера  $n_1 \times n_1$

Тогда свертка это операция над парой матриц  $M$  и  $Y_i$  в результате, которой получается матрица  $M'$  размера  $n'_1 \times n'_1$ , где  $n'_1 = n_1 - n_2 + 1$ , а элементы матрицы вычисляются вот так:

$$m'(i, j) = f \left( \sum_{t=1}^{n_2} \sum_{p=1}^{n_2} m(i+t-1)(j+p-1) \cdot Y(t, p) \right)$$

$f$  - по старой русской традиции это функция активации

Сколько ядер, столько и матриц  $M'$  в первом свертосном слое, затем в первом выборочном

слое мы строим матрицы  $M''$  (их будет столько же), затем из каждой  $M''$  снова получаем столько матриц, сколько ядер. Этот процесс продолжается пока не получатся матрицы из одного элемента, которые пойдут в нейроны выходного слоя, а там от них посчитают функцию активации и получится ответ.

Есть 2 дефолтные функции выборки:

-  $\mu(x) = \max(A)$

-  $\mu(x) = \text{avg}(A)$

где  $A$  это подматрица матрицы выборочного слоя размера  $n_3 \times n_3$ , где  $n_3$  обычно задает раз-  
работчик

## 6 На этом всё

всем спасибо, всем пока больше не приду не хочу допсу по английскому, ставьте лайки, любите маму, конспект скоро появится на моем гитхубе: <https://github.com/GeruniaSun/ITMO>