Нейронки. Лекция 1. Лисичка

Геруня солнышко

09.04.2024

1 Интродукшн

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

2 Вводные по нейроным сетям

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 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
- прогнозирование
- распознавание образа

Дефолтные архитектуры:
- перцептрон (от анг. perception - восприятие)
- сверточная нейронная сеть

Перцептрон := $G(X, U)$ - орграф 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 $X \backslash (X_{in} \cup X_{out})$ - 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: $\forall u \in U : w(u) \in [-1; 1]$ - ребро нулевого веса не оказывает влияния на сигнал, чем больше модуль веса, тем больше влияние этого ребра на сигнал(положительное или отрицательное)
Обозначим за $input(x), output(x)$ сумму весов входящих и выходящих стрелок для вершины x.
Заметим, что $\forall x \in X_{in} : \; input(x) = output(x)$, а
$\forall x \in X_j \in (X \backslash X_{in}) : \; input(x) = \sum\limits_{\forall x_i \in X_{j-1}} output(x_i) \cdot W_{ix}$ 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 $\begin{cases} f(a) = a, \ if \ a > 0 \\ f(a) = 0, \ if \ a \leq 0 \end{cases}$

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 мегаультра 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 подкрепением

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 нейроной сети требуется прежде всего критерий, который будет определять расстояние между правильными сигналами и полученными, то есть некоторая функция $E(W)$, где $W := (w_1, w_2, \ldots, w_{|U|})$ - множество весов стрелок сети

Замечание: $E(W) \in C^1[-1;1]$ другими словами функция критерия непрерывна и диференцируема на отрезке $[-1;1]$
$\sqsupset E^1(W),\ E^2(W),\ \dots\ ,\ E^k(W)$ - последовательность значений критерия на каждом шаге обучения ведь было бы славно, чтоб такая последовательность была убывающей
Тогда поймем, когда пора остановить обучение, возможности такие:
k_{max} то есть заданное наперед количество шагов обучения, или
$\left| E^k(W) - E^{k-1}(W) \right| \leq \varepsilon$ заданая наперед точность сети

$$W_{ij}^k = W_{ij}^{k-1} + \Delta W_{ij}^k \qquad (*)$$

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

$$\Delta W_{ij}^k = h^k \cdot p_{ij}^k \qquad (**)$$

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 минимума надо выбирать 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 врядли посещу.