Метод обратного распространения

Обозначения

1. Входы и выходы нейронной сети.

Пусть N – число наблюдений в обучающей выборке. (Предикторы и отклик вместе)
Каждое наблюдение – вектор $X_1, \ldots, X_N$.

Координаты наблюдений будем обозначать $X_i = \left(-1, x_{i1}, x_{i2}, \ldots x_{ip}\right)$. Напомним соглашение, в соответствии с которым первая координата каждого вектора равна -1 и появилась искусственным образом, она соответствует пороговому значению.

Из набора наблюдений образуем (потенциально) бесконечную последовательность.
Для этого N наблюдений случайным образом упорядочивают, они образуют первые N элементов последовательности, затем исходный набор заново упорядочивают и добавляют в конец очереди, тем самым получают элементы с N+1 го по 2N, и так далее.

Элемент последовательности наблюдений номер n будем обозначать $X(n)$.

Заметим, что вся последовательность оказалась разбита на части, по N нейронов в каждой. Состав элементов в каждой части один и тот же, каждая часть состоит из всех элементов выборки. Части различаются лишь порядком, в котором расположены наблюдения. Каждая такая часть будет у нас называться эпохой (в некоторых книгах эоном, изредка циклом). Если данных много, поступаем иначе (см. обсуждение minibatch'ей)

Для каждого элемента последовательности $X(n)$ известен желаемый выход, то есть вектор $O(n)$, который надо получить в результате работы обученной нейронной сети.

2. Обозначения элементов нейронной сети.

Для определенности рассмотрим нейронную сеть с одним входным слоем, двумя внутренними и одним выходным слоем. Обозначим слои нейронной сети буквами:
L – входной слой, A – первый внутренний слой, B – второй, C – третий, он же выходной слой сети. Пусть в каждом слое будет $m(L)$, $m(A)$, $m(B)$, $m(C)$ нейронов, соответственно.
Нейроны входного слоя L будем перечислять с помощью индекса l, $l = 1, 2, \ldots, m(L)$,
Нейроны входного слоя A будем перечислять с помощью индекса i, $i = 1, 2, \ldots, m(A)$,
Нейроны входного слоя B будем перечислять с помощью индекса j, $j = 1, 2, \ldots, m(B)$,
Нейроны входного слоя C будем перечислять с помощью индекса k, $k = 1, 2, \ldots, m(C)$,
Заметим, что вводя единообразие мы несколько увлеклись и продублировали обозначения, поскольку $p + 1 = m(L)$.

В начальный момент времени у нас имеется необученная нейронная сеть. В качестве начальных весов могут быть взяты случайные числа. Напомню, что процедура выбора начальных значений весов называется инициализация нейронной сети.

Веса выбираются случайным образом, но близкими к нулю.

Вопрос. Рассмотреть только что инициализированную, но еще не обученную нейронную сеть. Для определенности дополнительно предположим, что у нейронов внутренних слоев логистическая активационная функция. Такая сеть ведет себя примерно как ..?

Как линейная комбинация координат входных нейронов! (Почему?)

Веса этой нейронной сети будут изменяться в ходе обучения.

Выведем формулу, по которой будут вычисляться изменения.

Будем рассматривать ситуацию, когда нейронная сеть изменяется каждый раз после эксперимента с нейроном входной последовательности. Другой вариант, когда изменения весов нейронной сети происходят в конце каждой эпохи, реализуется сходным образом и рассматриваться не будет.

Рассмотрим вес какого-либо соединения ω. Этот вес меняется в ходе обучения нейронной сети. Начальное значение веса будем обозначать $\omega(0)$, значение веса после *n*-того шага обучения, то есть после эксперимента с *n*-м элементом входной последовательности, будем обозначать $\omega(n)$. Таким образом, входной последовательности наблюдений соответствует последовательность значений каждого веса.

Выше использовалось обозначение для произвольного веса соединения ω. Уточним обозначение, чтобы различать веса соединений разных нейронов.

Обозначим

$\omega_{li}^{LA}(n)$ вес соединения *l*-го нейрона входного слоя *L* и *i*-го нейрона слоя *A* после *n*-го шага обучения;

$\omega_{ij}^{AB}(n)$ вес соединения *i*-го нейрона входного слоя *A* и *j*-го нейрона слоя *B* после *n*-го шага обучения;

$\omega_{jk}^{BC}(n)$ вес соединения *j*-го нейрона входного слоя *B* и *k*-го нейрона слоя *C* после *n*-го шага обучения.

Конечно, это не самый экономный вариант обозначений… Но сейчас мы предпочтем ясность обозначений.

Напомним соглашение, в соответствии с которым $\omega_{0i}^{LA}(n)=\omega_{0j}^{AB}(n)=\omega_{0k}^{BC}(n)$ =-1. В соответствии с этим соглашением, данные веса соответствуют пороговым значениям.

Обозначим $y_i^A(n)$, $i=1,2,\dots,m(A)$, $y_j^B(n)$, $j=1,2,\dots,m(B)$, $y_k^C(n)$, $k=1,2,\dots,m(C)$ - выходные значения соответствующих нейронов.

Заметим, что

- выходные значения $y_i^A(n)$ и $y_j^B(n)$ являются в то же время входными значениями нейронов следующих слоев *B* и *C* соответственно.

- выходные значения нейронов внешнего слоя $y_k^C(n)$ являются одновременно и выходными значениями нейронной сети, на вход которой подан вектор $X(n)$.

Опишем действия, которые производятся в каждом нейроне, и введем необходимые обозначения. В каждом нейроне производится суммирование входных значений с соответствующими весами.

Обозначим эти линейные комбинации. Положим

$v_i^A=\sum_{l=0}^{m(L)} \omega_{li}^{LA}(n)\cdot x_l(n)$ - линейную комбинацию, которая считается в нейронах слоя *A*.

Аналогично,

$v_j^B=\sum_{i=0}^{m(A)} \omega_{ij}^{AB}(n)\cdot y_i^A(n)$ - линейная комбинация, которая считается в нейронах слоя *B*,

$v_k^C=\sum_{j=0}^{m(B)} \omega_{jk}^{BC}(n)\cdot y_j^B(n)$ - линейная комбинация, которая считается в нейронах слоя *C*.

Прежде, чем стать выходом нейрона, линейная комбинация преобразуется с помощью активационной функции. Будем рассматривать нейронную сеть, у которой во всех нейронах слоя используется одна и та же активационная функция. Случай разных активационных функций рассматривается аналогично. Будем обозначать активационные функции слоев $\varphi_A(t)$, $\varphi_B(t)$, $\varphi_C(t)$, соответственно.

Теперь можно выписать формулы для выходных значений нейронов. Имеем

$y_i(n) = \varphi_A(v_i(n))$, $\quad y_i(n)=\varphi_A(v_i(n)) \quad i = 1,2,\ldots,m(A)$

$y_j(n) = \varphi_B(v_j(n))$, $j = 1,2,\ldots,m(B)$,

$y_k(n) = \varphi_C(v_k(n))$, $k = 1,2,\ldots,m(C)$.

3. Ошибки нейронной сети.

Обсудим теперь ошибки, причем речь будет идти как о необученной нейронной сети, так и об обученной нейронной сети.

Начнем с обсуждения нейронов выходного слоя. На шаге обучения номер *n* на вход нейронной сети подается наблюдение $X(n)$, на выходе получаются значения $y(n) = (y_1(n), y_2(n),\ldots, y_{m(C)}(n))$, , а должны получится значения $o(n) = (o_1(n), o_2(n),\ldots, o_{m(C)}(n))$. Несовпадения $o(n)$ и $y(n)$ означают, что на шаге *n* нейронная сеть сделала ошибку $e(n) = o(n) - y(n)$. Критерием качества выберем квадратичное отклонение $E(n) = \frac{1}{2}\sum_{k=1}^{m(C)} e_k^2(n)$.

Так будет считаться ошибка для одного наблюдения. Чтобы измерять среднюю ошибку по всем наблюдениям, за одну эпоху, будем использовать среднеквадратичное отклонение $\frac{1}{N}\sum_{n=\tau N+1}^{\tau(N+1)} E(n)$, где τ - номер эпохи.

Идея метода обратного распространения.

Будем минимизировать среднеквадратичное отклонение методом скорейшего спуска.

…

Производная сложной функции

Упражнение. Найти производную функции $\sin^3(x^2)$

Рассмотрим нейрон выходного слоя под номером *k*. Пусть на вход нейронной сети был подан вектор $X(n)$, тогда на выходе нейрона получено значение $y_k(n) = \varphi_C(v_k(n))$, желаемое значение $o_k(n)$, а ошибка $e_k(n)$ получается вычитанием $e_k(n) = o_k(n) - y_k(n)$.

Наша цель – подправить веса входов нейрона $\omega_{jk}^{BC}(n)$, $j = 1,2,\ldots,m(B)$ так, чтобы уменьшить ошибку $E(n)$.

Метод скорейшего спуска предписывает следующую формулу для изменения весов

$$\omega_{jk}^{BC}(n+1) = \omega_{jk}^{BC}(n) - \gamma \cdot \frac{\partial E(n)}{\partial \omega_{jk}^{BC}(n)}$$

Будем искать производную $\frac{\partial E(n)}{\partial \omega_{jk}^{BC}(n)}$ как производную сложной функции. Получим

$$\frac{\partial E(n)}{\partial \omega_{jk}^{BC}(n)}=\frac{\partial E(n)}{\partial y_k^C(n)}\cdot\frac{\partial y_k^C(n)}{\partial v_k^C(n)}\cdot\frac{\partial v_k^C(n)}{\partial \omega_{jk}^{BC}(n)}$$

Поскольку

$$\frac{\partial v_k(n)}{\partial \omega_{jk}^{BC}(n)}=\frac{\partial}{\partial \omega_{jk}^{BC}(n)}\sum_{j=0}^{m(B)}\omega_{jk}^{BC}(n)\cdot y_j(n)=y_k(n);$$

$$\frac{\partial v_k^C(n)}{\partial \omega_{jk}^{BC}(n)}=\frac{\partial}{\partial \omega_{jk}^{BC}(n)}\sum_{j=0}^{m(B)}\omega_{jk}^{BC}(n)\cdot y_j^B(n)=y_j^B(n) \quad ;$$

$$\frac{\partial y_k^C(n)}{\partial v_k^C(n)}=\frac{\partial}{\partial v_k^C(n)}\varphi_C(v_k^C(n))=\varphi'_C(v_k^C(n)) \quad ;$$

$$\frac{\partial E(n)}{\partial y_k^C(n)}=\frac{\partial E(n)}{\partial e_k(n)}\cdot\frac{\partial e_k(n)}{\partial y_k^C(n)}=e_k(n)\cdot(-1)$$

получаем, что

$$\omega_{jk}^{BC}(n+1)=\omega_{jk}^{BC}(n)-\gamma\cdot y_k(n)\cdot\varphi'_C\left(v_k(n)\right)\cdot e_k(n)\cdot(-1)=$$
$$=\omega_{jk}^{BC}(n)+\gamma\cdot e_k(n)\cdot y_k(n)\cdot\varphi'_C\left(v_k(n)\right).$$

$$\omega_{jk}^{BC}(n+1)=\omega_{jk}^{BC}(n)-\gamma\cdot e_k(n)\cdot(-1)\cdot\varphi'_C(v_k^C(n))\cdot y_j^B(n) \quad =$$
$$= \quad \omega_{jk}^{BC}(n)+\gamma\cdot e_k(n)\cdot y_j^B(n)\cdot\varphi'_C(v_k^C(n))$$

Формула для изменения весов нейронов внешнего слоя получена.

Рассмотрим формулу для изменения весов нейронов внутреннего слоя.

Будем искать производную $\frac{\partial E(n)}{\partial \omega_{ij}^{AB}(n)}$ как производную сложной функции.

Получим

$$\frac{\partial E(n)}{\partial \omega_{ij}^{AB}(n)}=\frac{\partial E(n)}{\partial y_j(n)}\cdot\frac{\partial y_j(n)}{\partial v_j(n)}\cdot\frac{\partial v_j(n)}{\partial \omega_{ij}^{AB}(n)}$$

Поскольку

$$\frac{\partial v_j^B(n)}{\partial \omega_{ij}^{AB}(n)}=\frac{\partial\sum_i \omega_{ij}^{AB}(n)\cdot y_i^B(n)}{\partial \omega_{ij}^{AB}(n)}=y_i^B(n)$$

$$\frac{\partial y_j^B(n)}{\partial v_j^B(n)}=\frac{\partial}{\partial v_j^B(n)}\varphi_B(v_j^B(n))=\varphi'_B(v_j^B(n))$$

$$\frac{\partial E(n)}{\partial y_j^B(n)}=\frac{\partial(\frac{1}{2}\cdot\sum_k e_k^2(n))}{\partial y_j^B(n)}=\sum_k e_k(n)\cdot\frac{\partial e_k(n)}{\partial y_j^B(n)}$$

$$\frac{\partial e_k(n)}{\partial y_j(n)}=\frac{\partial e_k(n)}{\partial v_k^C(n)}\cdot\frac{\partial v_k^C(n)}{\partial y_j^B(n)}$$

$$\frac{\partial e_k(n)}{\partial v_k^C(n)} = \frac{\partial (o_k(n) - y_k(n))}{\partial v_k^C(n)} = -\frac{\partial y_k^C(n)}{\partial v_k^C(n)} = -\frac{\partial \varphi_C(v_k^C(n))}{v_k^C(n)} = -\varphi'_C(v_k^C(n))$$

$$\frac{\partial v_k^C(n)}{\partial y_j^B(n)} = \frac{\partial (\sum_j \omega_{jk}^{BC}(n) \cdot y_j^B(n))}{\partial y_j^B(n)} = \omega_{jk}^{BC}(n)$$

$$\frac{\partial E(n)}{\partial \omega_{ij}^{AB}(n)} = -\sum_k (e_k(n) \cdot \omega_{jk}^{BC}(n) \cdot \varphi'_C(v_k^C(n))) \cdot \varphi'_B(v_j^B(n)) \cdot y_i^B(n)$$

$$\omega_{ij}^{AB}(n+1) = \omega_{ij}^{AB}(n) - \gamma \cdot \frac{\partial E(n)}{\partial \omega_{ij}^{AB}(n)}$$