Омский Государственный ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

7. Регуляризация

Автор: Шевляков Артём Николаевич

Постановка проблемы

Проблема: большие по модулю веса

После обучения НС у нее могут оказаться **астрономически большие** (по абсолютной величине) веса.

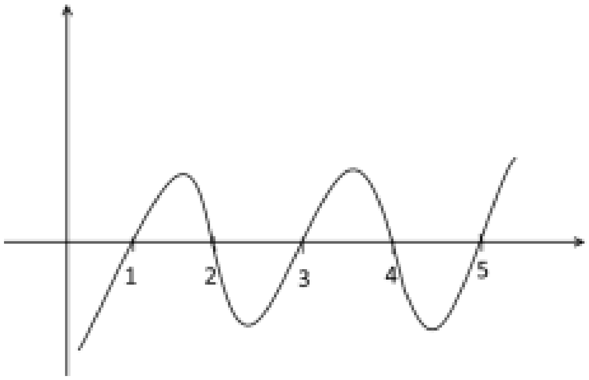
Почему это плохо?

- Когда у функции $F_{NN}(x)$ большие веса w_i, предсказание НС становится неустойчивым. Небольшое изменение в значениях нецелевых признаков приводит к большому изменению предсказанного признака.
- (философский) большие числа сами по себе менее вероятны для описания природы, чем небольшие числа.
- если задача предсказания допускает несколько решений, то нужно выбирать максимально простое (см. следующий слайд).

Наиболее простая зависимость

Рассмотрим задачу предсказания:

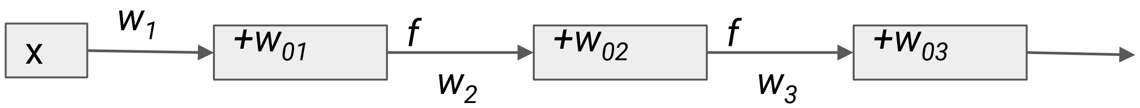
Наиболее естественное решение – это построение НС с функцией $F_{NN}(x)=0$, но неожиданно может натренироваться НС с функцией:



Номер курса, X	Самооценка студента, Y
1	0
2	0
3	0
4	0

НС со сложной зависимостью

Рассмотрим такую НС,



с ФА $f(x)=x^2$

$$F_{NN}(x) = w_3 f(w_2 f(x w_1 + w_{01}) + w_{02}) + w_{03},$$
$$F_{NN}(x) = w_3(w_2(x w_1 + w_{01})^2 + w_{02})^2 + w_{03}$$

Натренируем НС на данных:

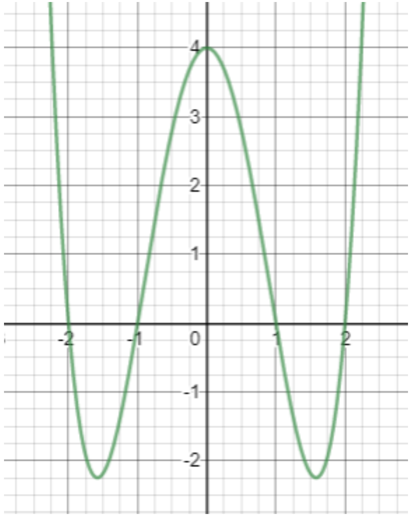
X	Y
-2	0
-1	0
1	0
2	0

НС со сложной зависимостью

ГС по функции потерь получит следующие веса:

$$w_1 = w_2 = w_3 = 1, w_{01} = 0, w_{02} = -5/2, w_{03} = -9/4,$$

$$F_{NN}(x) = (x^2 - 5/2)^2 - 9/4 = x^4 - 5x^2 + 4 = (x^2 - 1)(x^2 - 4)$$

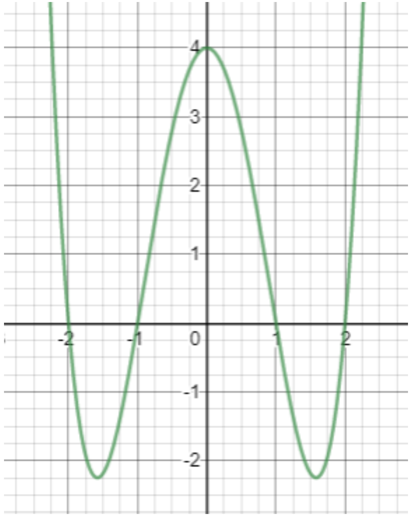


НС со сложной зависимостью

И мы получаем, что на ТВ НС дает абсолютно точные ответы.

Это произошло из-за того, что НС нашла слишком большие веса, по сравнению с нормальными (нулевыми) весами.

X	Y
-2	0
-1	0
1	0
2	0



Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

7. Регуляризация

Что делать?

Что делать?

Как запретить НС в процессе ГС получать слишком большие веса?

И главное не переусердствовать в этом! Возможно, в конкретной задаче большие веса действительно нужны.

1-й совет: **ГС нужно начать с маленьких весов.**
Для этого использовать:
а) **нормализацию данных**;
б) **инициализацию Ксавье**.

2-й совет: **применить эту самую** **регуляризацию**.

Регуляризация: неформально

Итак, нужно запретить НС искать слишком большие веса.

Это требование нужно **вписать в функцию потерь**. Но как?

Пусть $L(w)$ – «обычная» функция потерь (то есть сумма квадратов ошибок).

Мы приделаем ей «хвост» и получим регуляризированную функцию потерь:

$L_{reg}(w)=L(w)+C(w_1^2+...+w_k^2)$,

где C – некоторая **неотрицательная константа** (её надо самим выбрать), а в скобках стоит сумма квадратов всех весов (веса-связи и смещения) нашей НС. Затем мы как обычно ищем минимум функции $L_{reg}(w)$.

Смысл константы C

Константа *C* – это (простите за каламбур) вес весов нейронной сети.

Она отражает **компромисс** между двумя противоположными задачами:

- искать минимум «старой» функции потерь L(w);
- уменьшать веса сети.

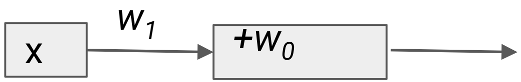
Чем больше *C*, тем выше приоритет второй задачи.

Чем ближе к нулю *C*, тем выше приоритет первой задачи.
В частности, при *C=0* мы получаем обычную процедуру тренировки НС.

Зависимость весов НС от константы С

Вспомним простенькую задачу. Натренировать НС для предсказания целевого признака Y по данным:

X	Y
-1	1
0	0
1	1
2	4



Мы вычислили, что оптимальные веса равны $w_1=w_0=1$.
В задаче минимизировалась функция потерь

$$L(w)=(-w_1+w_0-1)^2+(w_0-0)^2+(w_1+w_0-1)^2+(2w_1+w_0-4)^2.$$

Зависимость весов НС от константы C

Теперь будем минимизировать функцию потерь

$$L_{reg}(w)=L(w)+C(w_1^2+w_0^2)=(-w_1+w_0-1)^2+(w_0-0)^2+(w_1+w_0-1)^2+(2w_1+w_0-4)^2+\mathbf{C(w_0^2+w_1^2)}$$

с помощью ГС (эта функция достаточная простая, у нее ровно один минимум). Получаем таблицу:

Константа C	Оптимальные значения весов
0	$w_1=1,\ w_0=1$
50	$w_1=0.14,\ w_0=0.11$
100	$w_1=0.07,\ w_0=0.06$
1000	$w_1=0.008,\ w_0=0.006$

Зависимость весов НС от константы С

Какое значение константы С самое лучшее?

Об этом будет рассказано в теме «**Выбор оптимальных гиперпараметров**».

Небольшой спойлер:

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

7. Регуляризация

Второй вид регуляризации

Вот это поворот!

Оказывается есть и другой вид регуляризации.

Ранее мы изучали регуляризированную функцию потерь

$L_{reg}(w)=L(w)+C(w_1^2+...+w_k^2)$ – такой вид регуляризации называется **L2-регуляризацией**.

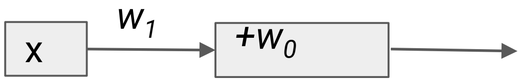
Но можно составить такую функцию потерь:
$L_{reg}(w)=L(w)+C(|w_1|+...+|w_k|)$ – это **L1-регуляризация.**

Посмотрим на примерах, в чем отличие
L1- и L2-регуляризации.

Зависимость весов НС от константы C

Опять помучаем нашу простенькую задачу.
Натренировать НС для предсказания целевого признака Y по данным:

X	Y
-1	1
0	0
1	1
2	4



«Обычная» функция потерь равна $L(w)=(-w_1+w_0-1)^2+(w_0-0)^2+(w_1+w_0-1)^2+(2w_1+w_0-4)^2$.

Зависимость весов НС от константы C

Теперь будем минимизировать функцию потерь

$L_{reg}(w)=L(w)+C(|w_1|+|w_0|)=(-w_1+w_0-1)^2+(w_0-0)^2+(w_1+w_0-1)^2+(2w_1+w_0-4)^2+\mathbf{C(|w_1|+|w_0|)}$

с помощью ГС.

Получаем таблицу:

Константа C	Оптимальные значения весов
0	$w_1=1,\ w_0=1$
1	$w_1=0.95,\ w_0=0.9$
5	$w_1=0.75,\ w_0=0.5$

Константа C	Оптимальные значения весов
9	$w_1=0.55,\ w_0=0.1$
10	$w_1=0.5,\ w_0=0$
15	$w_1=0.08,\ w_0=0$
20	$w_1=0,\ w_0=0$

Сравнение двух типов регуляризации

L1- и L2-регуляризации по разному уменьшают величины весов НС.

- в L2-регуляризации с ростом константы C веса **плавно стремятся к нулю**;
- в L1-регуляризации с ростом константы C все **веса постепенно зануляются**.

Какой из двух типов регуляризации выбрать?

Это тоже гиперпараметр, см. главу «**Выбор оптимальных гиперпараметров**».

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

7. Регуляризация

Выводы

Выводы:

- Мы поняли, что нужно бороться с бесконтрольным возрастанием абсолютных значений весов НС.
- Наиболее естественный способ борьбы – регуляризация.
- Есть два типа регуляризации, которые по-разному воздействуют на веса НС.