

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

10. Контроль качества обучения

Автор: Шевляков Артём Николаевич

Валидационная выборка

Как проверить точность НС

Вы натренировали НС. Как проверить ее точность и адекватность.

НС тренируется на ТВ и старается угадать ответы для объектов из ТВ.

Где гарантия, что для новых объектов (не из ТВ) НС будет вести себя адекватно?

Решение.

Вначале у вас есть большая **размеченная выборка (РВ)**, то есть множество объектов с известными ответами. РВ разбиваем на две части:

тренировочную (ТВ) и **валидационную** (ВВ) выборки в пропорции 80% и 20% (как правило).

Как и ранее, с помощью ТВ мы тренируем НС, а по ВВ мы проверяем точность и адекватность построенной НС.

Точность на ВВ

Формулы для точности НС существенно зависят от типа задачи:

- задача регрессии,
- задача классификации.

В любом случае все объекты ВВ нужно «пропустить» через НС и запомнить ответ, который выдает НС. Для объектов ВВ должна получиться таблица:

Объекты	Признак Y	Предсказанное значение
A_1	y_1	z_1
A_2	y_2	z_2
...
A_t	y_t	z_t

Точность задачи регрессии

Наиболее популярная метрика качества в задаче регрессии

$$MAE = \frac{1}{n} (|y_1 - z_1| + \dots + |y_t - z_t|)$$

Например, для таблицы MAE равен: $\frac{1}{4}(|1 - (-1)| + |0 - 0| + |1 - 1| + |4 - 2|) = 1$.

Объекты	Признак Y	Предсказанное значение
A ₁	1	-1
A ₂	0	0
A ₃	1	1
A ₄	4	2

С классификацией всё веселее

Прежде всего, информацию нужно занести в матрицу ошибок

Матрица ошибок		Истинный класс	
		0	1
Предсказанный класс	0	?	?
	1	?	?

Точность задачи классификации будет вычисляться по ячейкам матрицы ошибок.

С классификацией всё веселее

Объекты	Признак Y	Предсказанное значение
A ₁	0	0
A ₂	0	1
A ₃	0	0
A ₄	0	1
A ₅	1	0
A ₆	1	1
A ₇	1	1
A ₈	1	1

Матрица ошибок		Истинный класс	
		0	1
Предсказанный класс	0	2	1
	1	2	3

С классификацией всё веселее

Ячейки матрицы ошибок имеют собственные названия

Матрица ошибок		Истинный класс	
		0	1
Предсказанный класс	0	TN	FN
	1	FP	TP

TN – true negative

TP – true positive

FN – false negative

FP – false positive

Точность и полнота (precision and recall)

$$precision = \frac{TP}{TP + FP},$$

$$recall = \frac{TP}{TP + FN}$$

Это точность и полнота для класса 1.

Точность и полнота для класса 0 считаются аналогично.

Матрица ошибок		Истинный класс	
		0	1
Предсказанный класс	0	TN	FN
	1	FP	TP

Пример вычисления

$$\text{precision} = 3 / (3 + 2) = 0.6$$

$$\text{recall} = 3 / (3 + 1) = 0.75$$

Матрица ошибок		Истинный класс	
		0	1
Предсказанный класс	0	2	1
	1	2	3

$$\text{precision} = \frac{TP}{TP + FP},$$

$$\text{recall} = \frac{TP}{TP + FN}$$

Точность и полнота (precision and recall)

Идеально, чтобы точность и полнота были равны 1 (100%).

Но на практике увеличение одного показателя может приводить к ухудшению другого.

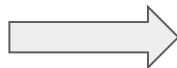
Матрица ошибок		Истинный класс	
		0	1
Предсказанный класс	0	TN	FN
	1	FP	TP

$$precision = \frac{TP}{TP + FP},$$

$$recall = \frac{TP}{TP + FN}$$

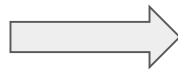
Крайние случаи (параноидальные классификаторы)

Матрица ошибок		Истинный класс	
		0	1
Предсказанный класс	0	2	1
	1	2	3



Матрица ошибок		Истинный класс	
		0	1
Предсказанный класс	0	0	0
	1	4	4

$$\text{precision} = 3 / (3 + 2) = 0.6$$



$$\text{precision} = 4 / (4 + 4) = 0.5$$

$$\text{recall} = 3 / (3 + 1) = 0.75$$

$$\text{recall} = 4 / (4 + 0) = 1$$

Одна характеристика вместо двух

Итак, качество классификации проверяется по двум характеристикам:
точность и полнота.

Это может приводить к появлению **несравнимых моделей.**

Например:

1ая НС, точность 0.9 полнота 0.8

2ая НС, точность 0.8 полнота 0.9

Как выбрать лучшую?

- мнение эксперта, какой параметр более важен для конкретной задачи;
- подсчитать агрегированную характеристику качества классификации.

F₁-значение (F₁-value)

F₁-значение позволяет вместо точности и полноты использовать только один параметр качества классификации.

Матрица ошибок		Истинный класс	
		0	1
Предсказанный класс	0	2	1
	1	2	3

$$F_1 = \frac{2 \cdot precision \cdot recall}{precision + recall}$$

precision=0.6, recall=0.75, $F_1 = 2 \cdot 0.6 \cdot 0.75 / (0.6 + 0.75) = 0.67$

Чем ближе F_1 к 1, тем лучше.

F_1 -значение (F_1 -value)

В задачах, в которых точность и полнота не равноценны применяют

$$F_{\beta} = \frac{(1 + \beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall} \quad (\beta > 0)$$

Параметр β показывает **значимость полноты по сравнению с точностью**.

Если $\beta = 1$, то мы получаем «обычное» значение F_1 .

F₁-значение (F₁-value)

Матрица ошибок		Истинный класс	
		0	1
Предсказанный класс	0	2	1
	1	2	3

$$F_{\beta} = \frac{(1 + \beta^2) \cdot precision \cdot recall}{\beta^2 \cdot precision + recall}$$

Было: precision=0.6, recall=0.75, $F_1 = 2 \cdot 0.6 \cdot 0.75 / (0.6 + 0.75) = 0.67$

Пусть: $\beta=2$, $F_2 = 5 \cdot 0.6 \cdot 0.75 / (4 \cdot 0.6 + 0.75) = 0.71$

Пусть: $\beta=3$, $F_3 = 10 \cdot 0.6 \cdot 0.75 / (9 \cdot 0.6 + 0.75) = 0.73$

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

10. Контроль качества обучения

Выбор оптимальных гиперпараметров

Что это такое?

18+

У нейросети и в алгоритмах её тренировки есть величины (параметры), которые сами не настраиваются, а задаются до процесса обучения человеком.



Дата-
саентист

Трехслойная НС

Что это такое?

У самой НС и в алгоритмах её тренировки есть величины (параметры), которые сами не настраиваются, а задаются до процесса обучения человеком.

Например:

- число слоев в НС, число нейронов в слое, итп (в общем, архитектура сети);
- шаг ГС h , константа C в регуляризации, итп (параметры обучения).

Такие величины называются **гиперпараметрами**.

Главная проблема: как задать оптимальное значение гиперпараметра?

Как найти оптимальное значение гиперпараметра

По сути с помощью перебора.

1. Формируем список возможных значений параметра $B=\{b_1, \dots, b_r\}$.
2. Для каждого значения b_i случайным образом делим PB на TB_i и VB_i .
3. Тренируем HC_i на TB_i и считаем её точность на VB_i .
4. Выбираем такое значение параметра b_i , на котором достигается максимальная точность.

Недостатки: на точность параметра b_i существенно влияет конкретное разбиение на TB и VB .

Значит, каждое значение параметра b_i нужно проверять на нескольких разбиениях TB и VB .

Кросс-валидация

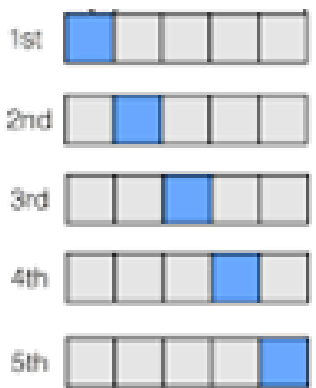
Разбиваем РВ на K фолдов (частей).

Каждое значение параметра b_i участвует в построении K штук НС:

$НС_{i1}, \dots, НС_{iK}$, причём каждый фолд ровно один раз является валидационным.

Для каждой $НС_{ij}$ считаем точность.

Итоговая точность для значения параметра b_i - это среднее арифметическое точностей K нейронных сетей $НС_{i1}, \dots, НС_{iK}$



Хорошее разбиение размеченной выборки

PB в идеале нужно разбить на **три** части: ТВ, ВВ и **тестовую** (ТеВ) выборки (стандартная пропорция 70% 10% 20%)

ТеВ в процессе тренировки и выбора оптимальных гиперпараметров **вообще не участвует**.

Когда будут выбраны оптимальные параметры, то итоговая НС проверяется на ТеВ. Полученная точность считается окончательной – ну а дальше решение за заказчиком.

Контроль шага обучения по ВВ

После каждой эпохи обучения вычислять точность НС на ВВ.
Если ошибка на ВВ перестала уменьшаться, то уменьшить шаг h .

Можно вообще **остановить обучение**, когда ошибка на ВВ начала расти.

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

10. Контроль качества обучения

Переобучение и недообучение

Самое страшное, что может случиться...

Если ошибка НС даже на ТВ высокая, то налицо **недообучение (underfitting)**.

Что делать:

нужно взять для тренировки более мощную нейросеть (с большим числом нейронов).

Самое страшное, что может случиться...

Если ошибка НС на ТВ низкая, но ошибка на ВВ высокая, то возникает **переобучение (overfitting)**.



Переобучение за счет
государства для безработных,
пенсионеров и беременных

ЮРИСТ МАКСИМ ПЕНИЗЕВ



Самое страшное, что может случиться...

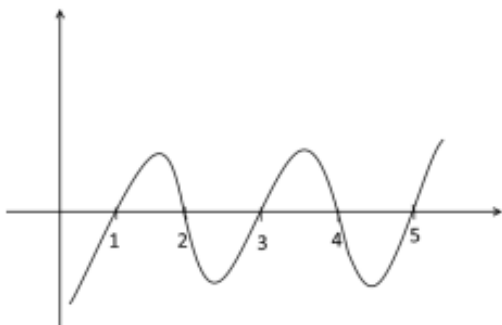
Если ошибка НС на ТВ низкая, но ошибка на ВВ высокая, то возникает **переобучение (overfitting)**.

В чем причина?

- НС просто запомнила всю ТВ (это как студент, который просто запомнил примеры из учебника, а закономерность не понял);
- НС реализовала неправильную закономерность (тут можно вспомнить старый пример, см. след. слайд)

Самое страшное, что может случиться...

НС восстановила закономерность:



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

А если ВВ состоит из «промежуточных» точек $x=1.5, 2.5, 3.5$, то ошибка на ВВ огромная!

Самое страшное, что может случиться...

Что делать?

- регуляризация,
- дропаут,
- больше данных.

Онлайн-курс

ВВЕДЕНИЕ В ИСКУССТВЕННЫЕ НЕЙРОННЫЕ СЕТИ

10. Контроль качества обучения

Выводы

Выводы:

- Мы рассмотрели разделение всей размеченной выборки на тренировочную, валидационную и тестовую.
- Мы привели характеристики, по которым вычисляется точность НС для задачи регрессии и для задачи классификации.
- Мы рассмотрели процедуру кросс-валидации, которая используется для оптимального выбора гиперпараметров.
- Рассмотрели явления недообучения и переобучения.