

Алгоритм решения задачи с Kaggle

Ипатов Дмитрий

24 декабря 2020 г.

Содержание

1 Введение	1
2 Выбранная задача	1
3 Входные переменные	1
4 Выходные переменные	2
5 Разбиение выборки	2
6 Нейронная сеть	2
7 Обучение	2
8 График обучения	2
9 Пример предсказания тестовой выборки	3
10 Вывод	3

1 Введение

Необходимо решить задачу с сайта [kaggle.com](https://www.kaggle.com) с использованием нейронной сети

2 Выбранная задача

В качестве задачи была выбрана задача - **Kepler Exoplanet Search**. Ссылка на задачу - <https://www.kaggle.com/nasa/kepler-exoplanet-search-results>

3 Входные переменные

В качестве входных переменных были выбраны все количественные значения из таблицы, за исключением *kepid*. Однако, в данных значениях слишком много null значений. Заполним их все средними значениями по столбцам соответствующим.

4 Выходные переменные

Целевой переменной в данной задаче является поле **koi-score**. Данное значение изменяется от 0 до 1. Поэтому очевидно, что в конце потребуется добавить сигмоидную функцию

5 Разбиение выборки

Давайте разобьем сначала по ключевой переменной на две выборки: ту, на которой далее будет обучаться (в ней все значения целевой переменной не null), и ту, у которой значения целевой переменной null, а далее попробуем предсказать эти значения. Далее выборку для обучения разобьем в соотношении 0.8 на тренировочную и 0.2 на валидационную выборки.

6 Нейронная сеть

В качестве библиотеки для создания нейронных сетей будем использовать keras. Зададим следующие характеристики ИНС:

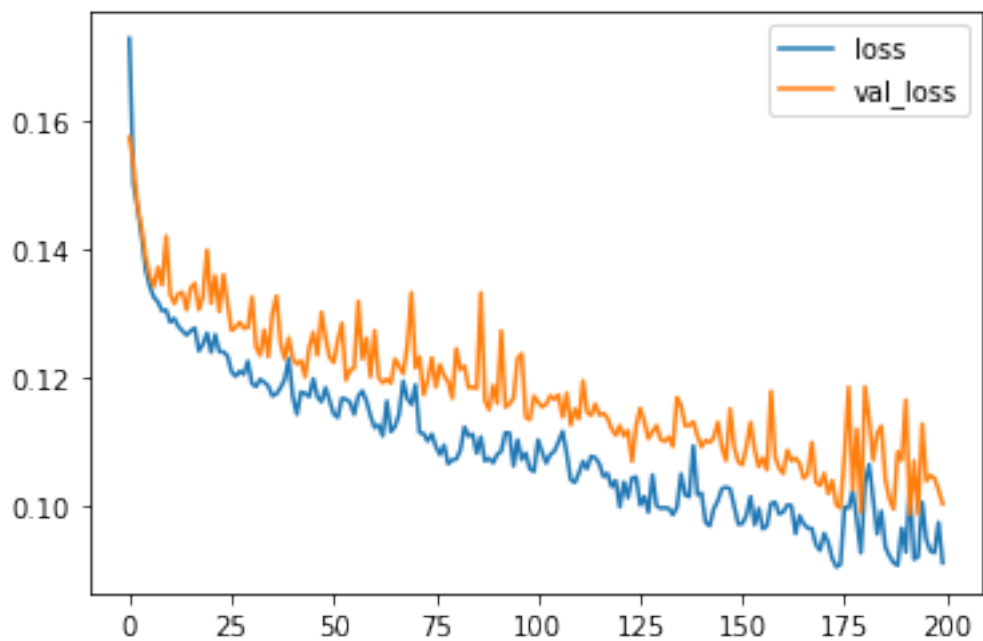
- на вход приходит 40 входных переменных;
- на первом слое содержится 128 нейронов с tanh функцией активации;
- на втором слое содержится 128 нейронов с tanh функцией активации;
- на третьем слое содержится 64 нейрона с sigmoid функцией активации;
- на выходное слое содержится 1 нейрон с сигмоидной функцией (так как целевая переменная изменяется от 0 до 1);

7 Обучение

В качестве метода оптимизации был выбран adam, а в качестве функционала ошибки - mae. Далее было проведено обучение на 200 эпохах с 256 батчами.

8 График обучения

На графике внизу показано изменение ошибки на тренировочной и валидационных выборках.



9 Пример предсказания тестовой выборки

```
[253]: test_y
```

```
Out[253]: array([[0.07052004],
 [0.6911719 ],
 [0.8923579 ],
 ...,
 [0.19576901],
 [0.35151732],
 [0.0492003 ]], dtype=float32)
```

10 Вывод

В результате данной работы была написана и обучена нейронная сеть с использованием библиотеки `keras` для предсказания величины `koi_score`, а также построены графики обучения