实验报告

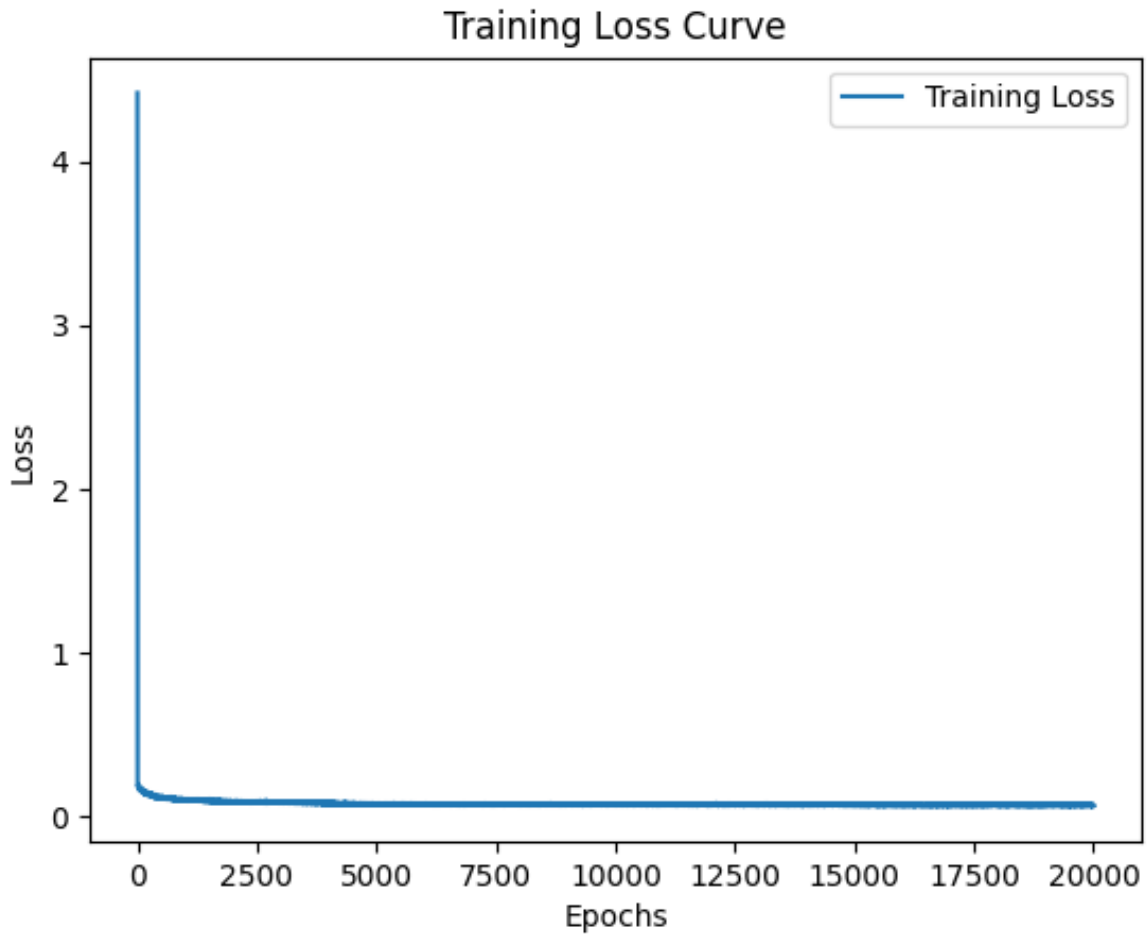
思路

设计了一个深度神经网络来预测房价，由于数据量其实比较少，只有几千条，所以神经网络不要太深，很容易训练不开。

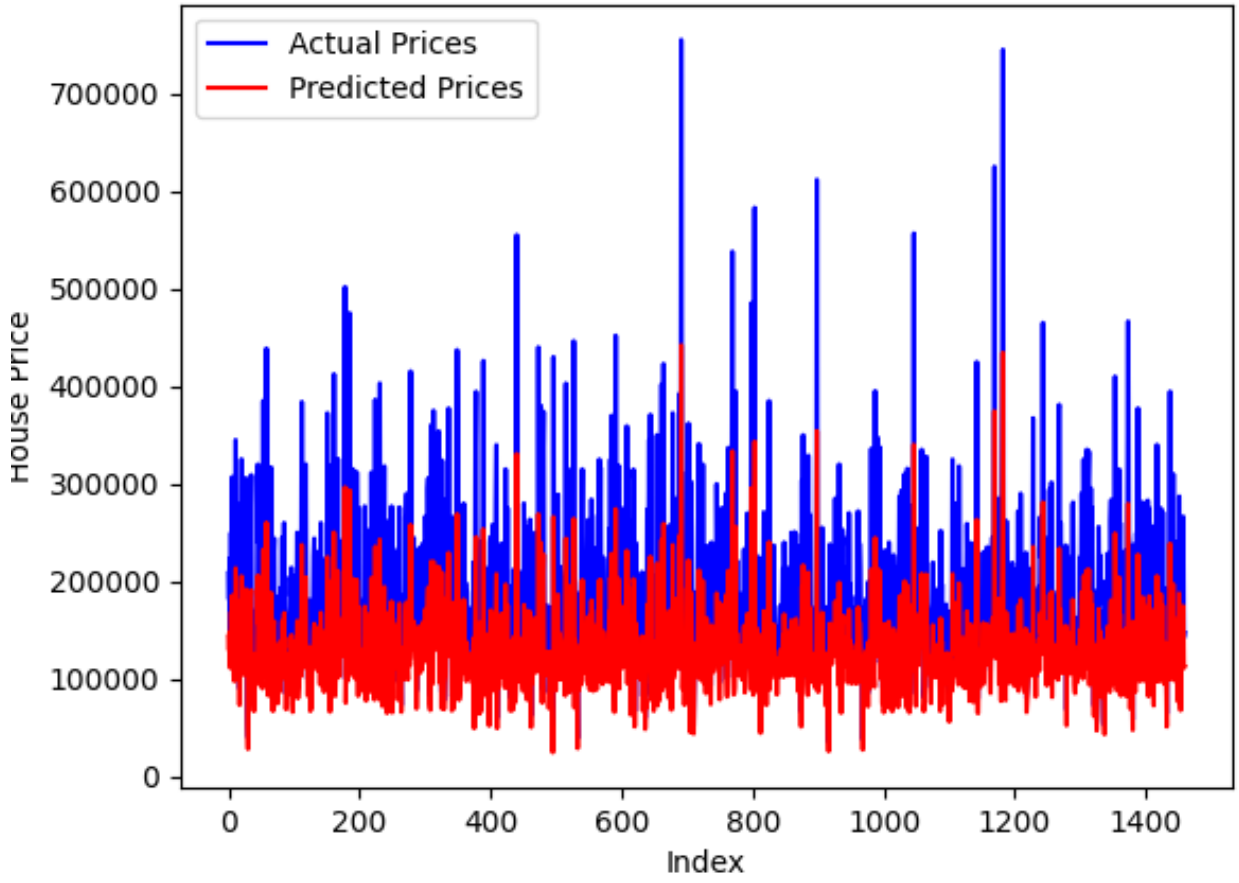
一开始使用 MSE 来作为损失函数，但 MSE 所产生的损失太大，非常容易导致训练不稳定、梯度爆炸等等问题。

后来使用 RMSE作为损失函数，即标签价格和预测价格都先加一再取自然对数，然后相减再开根号作为损失值。这样做的目的在于可以控制损失值在小区间内，避免出现训练不稳定、梯度爆炸的问题。这么做也有缺点，由于自然对数函数会压缩高输入值，所以使得房价的预测值几乎都偏低于真实值。

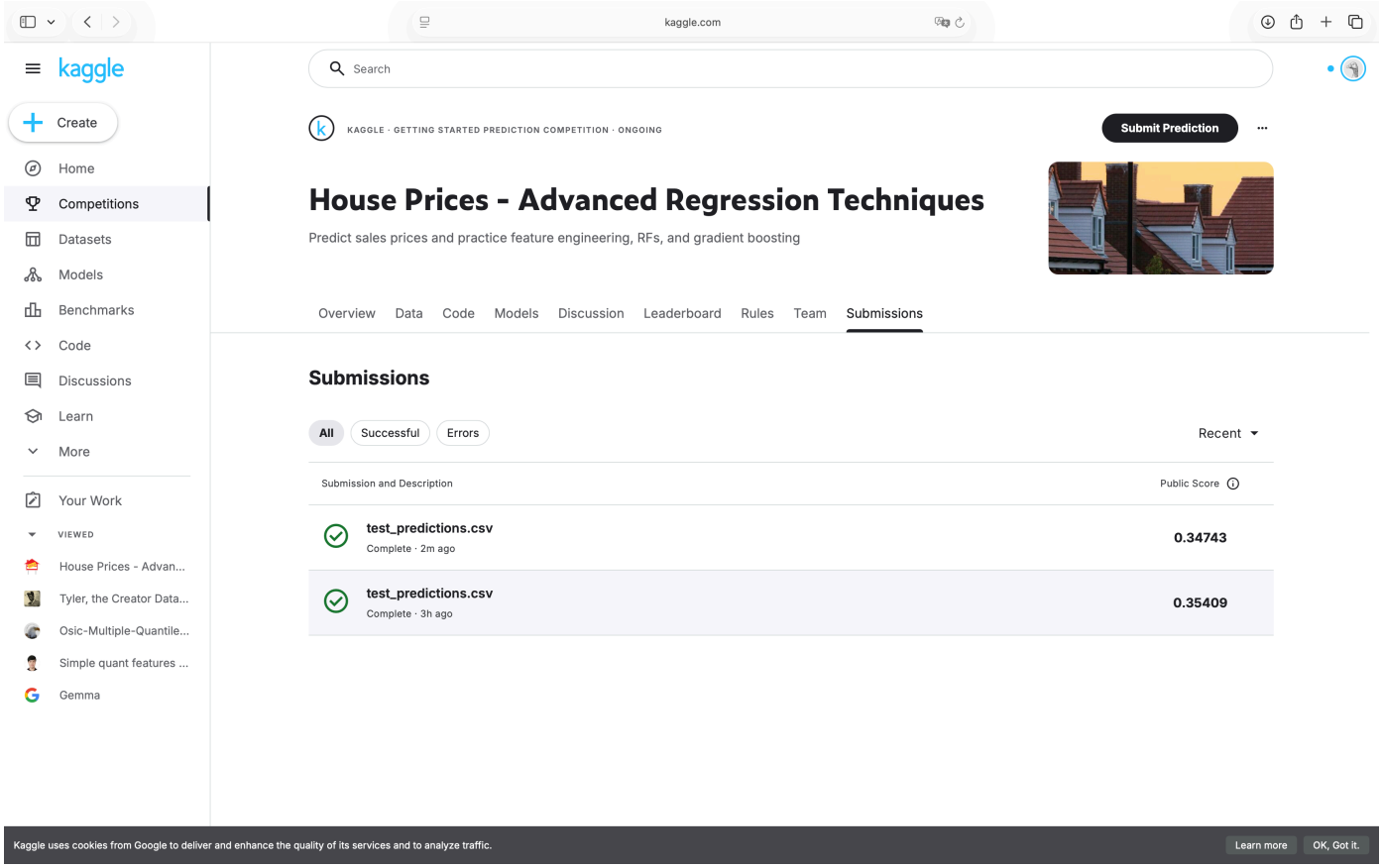
使用 RMSE 函数训练了 2000 个 epoch，稳定下降。

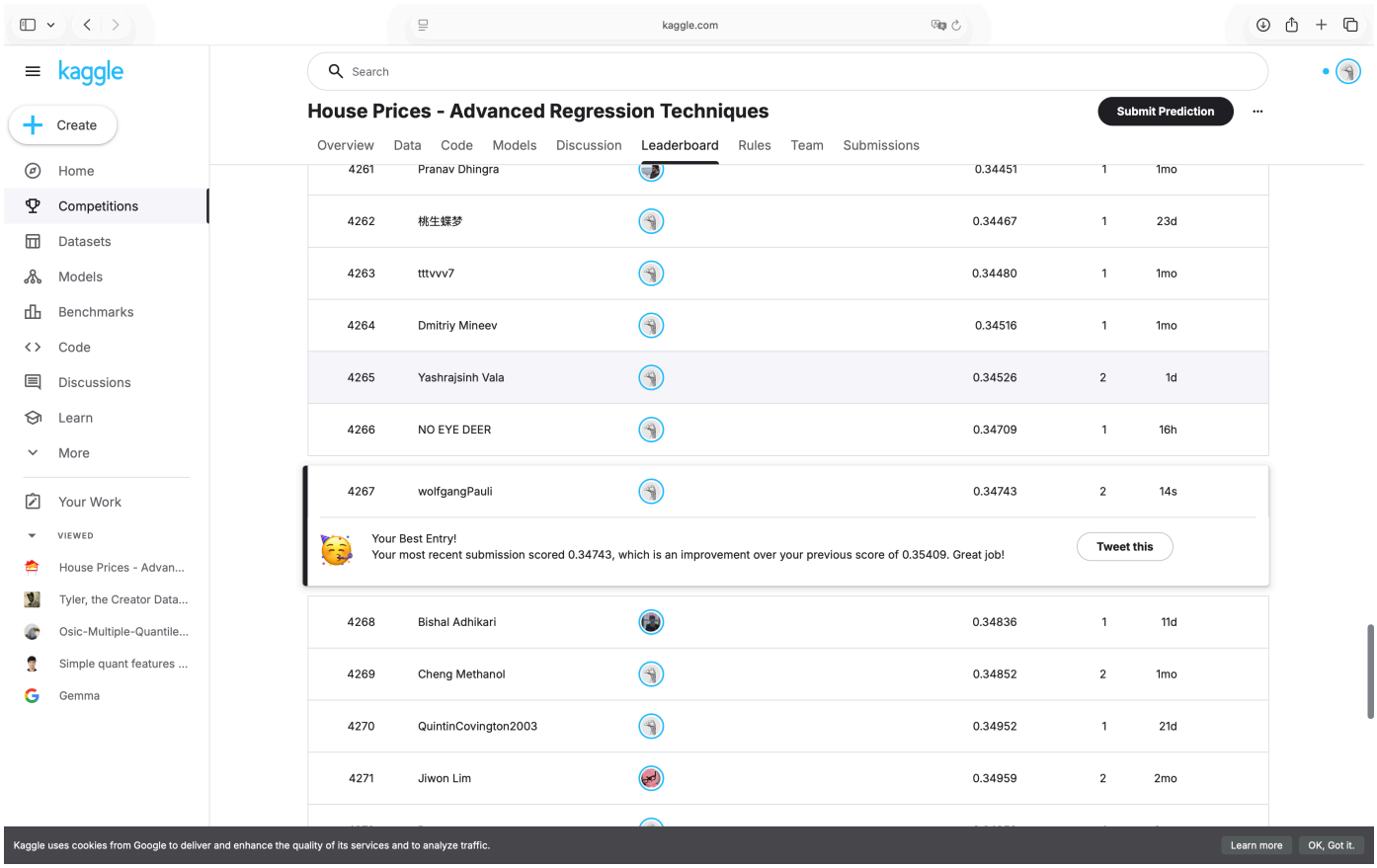


最后的验证结果比对如下。


Actual Prices
Predicted Prices
House Price
700000
600000
500000
400000
300000
200000
100000
0
0
200
400
600
800
1000
1200
1400
Index

提交截图


kaggle.com
kaggle
Search
Create
Home
Competitions
Datasets
Models
Benchmarks
Code
Discussions
Learn
More
Your Work
VIEWED
House Prices - Advan...
Tyler, the Creator Data...
Osic-Multiple-Quantile...
Simple quant features ...
Gemma
KAGGLE · GETTING STARTED PREDICTION COMPETITION · ONGOING
Submit Prediction
House Prices - Advanced Regression Techniques
Predict sales prices and practice feature engineering, RFs, and gradient boosting
Overview
Data
Code
Models
Discussion
Leaderboard
Rules
Team
Submissions
Submissions
All
Successful
Errors
Recent
Submission and Description
Public Score
test_predictions.csv
Complete · 2m ago
0.34743
test_predictions.csv
Complete · 3h ago
0.35409
Kaggle uses cookies from Google to deliver and enhance the quality of its services and to analyze traffic.
Learn more
OK, Got it.



收获与反思

1. 应根据不同的数据分布使用不同的损失函数

1. 将字符串类型的数据转换为分类型数据
2. 模型不应太大，因为数据量有限，容易训练不开