Assigment 1 report

Vladimir Klepikov
v.klepikov@innopolis.university

1 Motivation

My task about prediction different types of game streaming parameters. This field of work very interesting in last time. More and more people have simple computer with good internet connection and they wand play something vert heavy. So, we need quality service for users. It require monitoring of multiple parameters and setting more suitable parameters of game quality to make gameplay maximally smooth

2 Data

In our dataset we have some parameters, that describe stream parameters and user feedback on it

1. **RTT:** Round-trip parameter is time from computer to server to computer. Lower is better, it helps to reduce input lag.
2. **Dropped Frames:** Lost video frames show connection stability and some PC problems. Lower is better.
3. **FPS:** Amount of showing frames per second. It has more preferable amount, multiple by 30, like 30 or 60 FPS, due to some human's eye features and bugs.
4. **Adaptive bitrate:** Some categorical feature, that may be related with connection stability. It not needed when connection stable, but impotant when it fluctuate.
5. **Auto FEC:** Automatic Forward Error Correction, it also calls lag-compensate. Need on big ping connection.
6. **Targets:** Stream(classification) and bitrate(regression) prediction targets

3 Exploratory data analysis

This dataset have categorical and parametric features. If regression task we don't need categorical features, but in classification we have it. In simple way we can just use OneHotEncoder, but generally speaking, it generate big amount of zeroes, that not very suitable for next doings. So, column autobitratestate was created via pandas profiler. After this move I separate features and target and normalize it via MinMaxScaler. For regression task I build correlation matrix, look at parameters and decide to chop out 1 percent of minimal and maximal data. And, with this matrix i see that very big correlation of target with bitratestd, because it is literally same thing, so i don't use this parameters

4 Task

4.1 Regression

For solution of this task I use:

1. Lasso Linear Regression
2. Ridge Linear Regression
3. Simple Polynomial Regression

Lasso should show something like Ridge or worse, so it here just for fun

4.2 Classification

For the classification task, I have chosen Logistic regression algorithm with L1 and L2 regularization. Logistic regression is an extension of the linear regression algorithm for classification. This approach separates the classes with sigmoid-like function.

5 Results

This results looks bad, especially if looks at MSE error. But if dive deeper, we can see, that regression dataset have target in approx. [0,30000] after removing 1 percent of strange data. So, average error is about 15 percent. This is big parameter, but we can later work with this data using another instruments. Let's describe Regression scores

Table 1. Regression scores

Model	Lasso	Ridge	Polynomial
MSE	34302951	34302951	34508338
MAE	4505.402	4505.4	4527
R2Score	0.03807	0.0380734	0.032313

Table 2. Classification scores

Score	LogReg L1	LogReg L2	SVC
Accuracy	0.940533	0.940808	0.938287
Precision	0.708708	0.726788	0.979681
Recall	0.130687	0.130177	0.043010

6 Conclusion

So, it was very interesting task, as for me. I don't know how to use it in business cases, because it not very strong or precise, but like something learning or scientific it looks well.