FDA-hw3-2

Chosen Dataset & Problem Definition

Dota2 games Results Data Set
This dataset gives us some information when the game is running, and I'll predict the game result (which team wins) of the Dota2 computer game.

Analyze the data

We have some fields in the dataset, namely

- Result
- Cluster ID (a.k.a location ID)
- Game mode
- Game type
- hero indicator

I would like to take a look at the Game mode and Game type. Inspect if there exists human-recognizable difference distribution.

The result is as follows, it seems that they have simliar distribution.

XXXXFIGUREXXX

After that, I one-hot encoded the features, so that the table now looks like below.

Results

Classifiers from scikit-learn

I first tried the classifiers from scikit-learn, e.g. random forest classifier and logistic regression. Both of them got a bad result (approximately 52~53% in accuracy). I think this is because of the complexity of the model, so later on I tried neural network to train my model.

In neural network, I got slightly better result (which is 59.xx%~60% in accuracy). However, no matter how I tune the hyperparameter, the accuracy seem not to be over 60%. I think this is because of the limited number of features. I cannot find a way to expand the dota2 dataset, since the given features are so simple to compose new one.