README for the performance measurement part

Group 2

December, 2nd, 2019

Abstract

This Chapter is describes the evaluation of the best sampling and model combination, which is computed in the script 'PerformanceMeasurement.R'.

The goal of our performance measurement

After training a model it is important to check how well it performes. Since we not only trained one model but 140 in total (7 algos * 5 datasets * 4 positions), it is even more important to compare all the different models to see which one is the best. For our business case, this means that we want to see which combination of method and sampling we should take to predict which position, or if we even have a model, which does it better when we leave them together.

How to compare the method/sampling combinations for all positions

After training the models with 10-fold cross-validation, we applied them on all the (unsampled) data from 2007 to 2013 to obtain the true positives, true negatives, false positives and false negatives. Since we used the cross-validation for training, this will not give us the training error, but the testing error we get with all the available data prior to the 2014 NFL Draft. At the end of every model-script, we save this information separately, to bring it all together in the script 'PerformanceMeasurement.R'.

The first step, is to bring them all into one dataframe, to use them easier. Then we make sure, that we used the same, unsampled data by computing the sums of TP, TN, FP, FN, for every method/sampling/position combination. In the CheckTibble we now see, that for every position the whole column contains always the same number, which means this is the case.

Then we calculate the ratio of correct classification, which is equal to:

$$\frac{CorrectClassifications}{AllClassifications} = \frac{TP+TN}{TP+TP+FP+FN}$$

The result is a table showing the percentage of correct classification for all the 140 model/sampling/position combination. This table is quite big, but it is still interesting to have a look at it, to see how well which combination performes.

Table 1: Ratio of correct classification

Method	Sampling	QB	WR	RB	Together
ClassificationTree	no_sampling	0.9098	0.9208	0.9156	0.8939
ClassificationTree	oversampling	0.8780	0.8415	0.8631	0.7930
ClassificationTree	undersampling	0.7098	0.8254	0.8137	0.7938
ClassificationTree	Rose_both	0.8317	0.7985	0.8057	0.7746
ClassificationTree	Smote	0.8439	0.8062	0.7850	0.7998
KNN	no_sampling	0.8488	0.8892	0.8646	0.8755
KNN	oversampling	0.8366	0.8915	0.8519	0.8725
KNN	undersampling	0.6463	0.8331	0.7962	0.7994
KNN	Rose_both	0.7317	0.8192	0.7834	0.7981
KNN	Smote	0.8610	0.9023	0.8726	0.8918
NaiveBayes	no_sampling	0.8268	0.8908	0.8758	0.8901
NaiveBayes	oversampling	0.7000	0.7746	0.8137	0.7998
NaiveBayes	undersampling	0.6878	0.7946	0.8328	0.7968

Method	Sampling	QB	WR	RB	Together
NaiveBayes	Rose_both	0.7341	0.7231	0.8615	0.7930
NaiveBayes	Smote	0.7317	0.8077	0.7213	0.8798
randomForest	no_sampling	0.6829	0.7838	0.7452	0.7626
randomForest	oversampling	0.5192	0.5146	0.5009	0.4980
randomForest	undersampling	0.5254	0.4665	0.4607	0.4786
randomForest	Rose_both	0.5077	0.5139	0.4750	0.4970
randomForest	Smote	0.5389	0.5161	0.5136	0.5058
ANN	no_sampling	0.8488	0.9054	0.8838	0.8828
ANN	oversampling	0.8593	0.8664	0.8574	0.8531
ANN	undersampling	0.8559	0.8771	0.8708	0.8440
ANN	Rose_both	0.8923	0.8570	0.8558	0.8512
ANN	Smote	0.8639	0.8853	0.8692	0.8628
LogisticRegression	no_sampling	0.8510	0.9072	0.8938	0.8980
LogisticRegression	oversampling	0.7876	0.8608	0.8246	0.8432
LogisticRegression	undersampling	0.7678	0.8637	0.8167	0.8432
LogisticRegression	Rose_both	0.7879	0.8606	0.8248	0.8451
LogisticRegression	Smote	NA	NA	NA	NA