GE 461 – Data Stream Mining Assignment

Batıhan Akça
Industrial Engineering
Bilkent University
Ankara, Çankaya, Turkey, akcabatihan@gmail.com

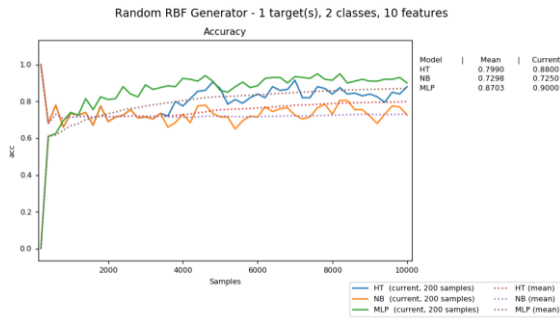## Separate Online Single Classifiers
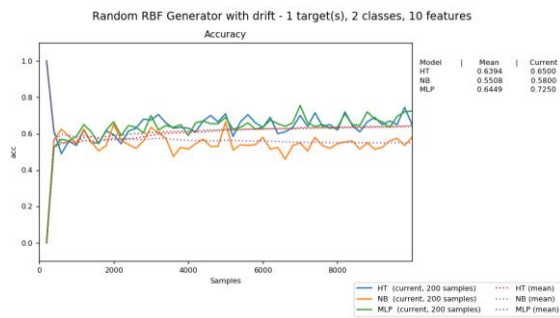

Figure 1
Single Classifiers for RBF Data Stream

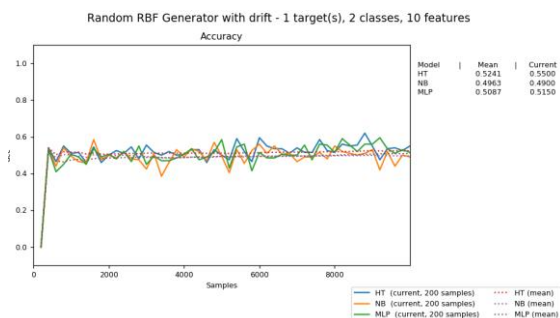*Figure 2*
Single Classifiers for RBF10 Data Stream

Figure 3
Single Classifiers for RBF70 Data Stream

Ensemble Online Classifiers

	APPROACH	ACCURACY
RBF	MV	0.884
RBF 10	MV	0.6177
RBF 70	MV	0.5932
RBF	WMV	0.8913
RBF 10	WMV	0.6146
RBF 70	WMV	0.478

Separate Batch Classifiers

HT	0.849
NB	0.706
MLP	0.949

RBF Dataset Accuracies

HT	0.65
NB	0.5185
MLP	0.765

RBF10 Dataset Accuracies

HT	0.542
NB	0.5105
MLP	0.5435

RBF70 Dataset Accuracies

Batch Ensemble Classifiers

	APPROACH	ACCURACY
RBF	MV	0.846
RBF 10	MV	0.549
RBF 70	MV	0.5525
RBF	WMV	0.8465
RBF 10	WMV	0.547
RBF 70	WMV	0.544

For the RBF data, MLP gives the best accuracy with 0.87 (Figure 1) among the single online classifiers where both of the online classifier methods overperform compared to MLP with 0.884 for MV method and 0.8913 for WVM.

For the RBF 10 data, MLP gives the best accuracy but the result is not as good as for the RBF data and the accuracy score is 0.6449 which is better than both of the online classifiers. Online classifies results are around 0.61 for both of the methods.

For the RBF 70 data, HT gives the best accuracy with 0.5241 among the single online classifiers. WMV method performs worse than every single online classifier but MV method beats HT accuracy by 0.5932.

Except RBF 10 data, ensemble method improves the accuracy most of the time and gives better results than single online classifiers.

If batch classifiers are examined, in overall results are worse than the online classifiers and in comparison, with batch classifiers and their ensemble methods, single batch classifiers overperform the ensemble methods and for the two MV and WMV methods, its ensemble methods gives very similar results.

For RBF data MLP single batch method gives 0.949 accuracy which is the highest among every instance of the assignment. Its batch ensemble method gives 0.846 for MV method and 0.8465 for WMV method.

For RBF10 data again the MLP gives the best accuracy with 0.765 with a huge difference between the other two single batch methods, 0.65 for HT and 0.5185 FOR NB. The ensemble methods for this data do not improve very much, two of the methods gives accuracy score around 0.55.

For RBF70 data any of the methods reach pass above the accuracy 0.55. It may be claimed that for such a data with a huge drift in it, online ensemble methods can give better results.

In overall the examination shows that ensemble methods increase the prediction accuracy for online learning methods, and they are better for online learning. For batch classifiers this is not the case most of the time.

REFERENCES

Pedregosa, F., Varoquaux, Ga"el, Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... others. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research, 12*(Oct), 2825–2830.