Lab Report
Massive Online Analysis (MOA) Lab

ZHANG Xin

January 2020

1 Classifiers selected for the experiments

To the purpose of covering more types of classifiers, seven classifiers are chosen:

1. **Majority Class Classifier** the majority classier is simple and direct, it is used as a baseline here for other comparisons.
2. **Drift Detection** handling concept drift datasets with a wrapper on a classifier
3. **Hoeffding Adaptive Tree** To represent the classification tree solution
4. **Leveraging Bagging** Leveraging Bagging for evolving data streams using ADWIN, to be compared with OzaBagADWIN
5. **Multinomial Naive Bayes classifier** As it uses a multinomial distribution for each of the features for a Naive Bayes classifier, more suitable to represent the Naive Bayes Method
6. **OzaBagADWIN** Bagging for evolving data streams using ADWIN.
7. **SAMKNN** Self Adjusting Memory (SAM) coupled with the k Nearest Neighbor classifier (kNN)

2 Results of the experiments

The results of the experiments can be found in the table below:

Table 1: Results

Classifier	Evaluation Time	Classification Correctness	Kappa Statistic
Majority Class	3.69	48.76	1.41
Drift Detection	7.92	88.03	80.78
Hoeffding Adaptive Tree	11.79	81.90	71.25
Multinomial Naive Bayes	5.58	62.46	37.69
Leveraging Bagging	86.14	91.70	86.67
OzaBagADWIN	56.31	84.74	75.63
SAMKNN	548.91	92.92	88.60

3 Discussion about the results

We can see that as a Baseline, the **Majority Class** is the fastest, but the overall correctness is below 50%. Moreover, the kappa statistic is only at 1.41%, whichi means the model is not reliable.
Next we can see the Multinomial Naive Bayes takes sightly more time to train, but the correctness and kappa score is still not ideal.
All other options achieved a correctness above 80%, it can be seen that drift detection is better than hoeffding adaptive tree and OzaBagADWIN in all means in terms of time, correctness and kappa score.
the best results (above 90%) are delivered by Leveraging Bagging and SAMKNN. But the training of SAMKNN takes much more time

4 Classifier Recommendation

So the model that I recommend to use with the CoverType dataset is **Leveraging Bagging**. We can see that it delivers a good precision, and the kappa statistic also indicates that the model trained is reliable. Even though that the SAMKNN classifier gave a better result and a more reliable model, it takes several times more time to train the model, but the improve of result is relatively insignificant.