MACHINE LEARNING FOR DATA STREAMS

BAYESIAN NETWORKS

NAÏVE BAYES

- See "A Survey on Supervised Classification on Data Streams" for information about naïve Bayes for data streams.
    - ➢ The naïve Bayes approach is intrinsically incremental and can be easily updated on-line. To do so, **it is sufficient to update some counts in order to estimate the univariate conditional probabilities**. These probabilities are based on an estimate of the densities; this problem can be viewed as **the incremental estimation of the conditional densities**
    - ➢ The naive Bayes classifier uses conditional probabilities estimates. **These estimates are usually done after a discretization of the explicative variables**. In the stream mining context this first step needs dedicated methods as most off-line methods usually need to load all the data in memory. In the literature, two kinds of publication related to incremental discretization can be found. **The articles related to data-mining are not numerous** but the literature related to database provides much more articles because the database systems (DBMS) need to have good estimates of the data distribution.

1. **Incremental Weighted Naive Bayes Classifiers (2014):**
    - o Presents a new method based on a graphical model close to a neural network which computes the weights on the input variables using a stochastic estimation. The method is incremental and produces an Weighted Naive Bayes Classifier for data stream. This weighting produces a single model close to an \averaged naive Bayes classifier" a "weighted naive Bayes classifier".
    - o Its complexity to predict is very low which makes it suitable and widely used for stream mining prediction (only depends on the number of explanatory variables).
    - o Its memory consumption is also low since it requires only one conditional probability density estimation per variable.
    - o The method used to update the weights is the one used usually for a standard back-propagation and three main parameters have to be considered (Lecun et al. (1998)) in the case of stochastic gradient descent: (i) the cost function; (ii) the number of iterations; (iii) the learning rate.
    - o The conditional density probabilites used as inputs on their graphical model are estimated using three methods: their two layers incremental discretization method based on order statistics as described in (Salperwyck and Lemaire (2013)) (ii) a two layer discretization method

"cPiD" which is a modified version of the PiD method of Gama (Gama and Pinto (2006)) (iii) and a Gaussian approximation.

- o Their WNB approach uses a low amount of memory thanks to the two level approach to estimate the conditional densities, and is purely incremental thanks to the graphical model and the stochastic gradient descent to estimate the weights.

- o Comparison (Weighted Naive Bayes trained online with the two level discretization method which uses GKClass (level 1) and the MODL discretization (level 2) and our method based on the graphical method to estimate the weights.)with naive Bayes (NB) trained offline with the MODL discretization and all the data in memory; (2) an Averaged Naive Bayes (ANB) trained offline with the MODL discretization and a Naive Bayes trained online with the two level discretization method which uses GKClass (level 1) and the MODL discretization (level 2):
    - It improves the performance compared to the non-weighted version and is close to the off-line averaged version of the naive Bayes classifier.
    - They think they can improve the results in future work:
        - Use the GK Class summaries as "mini-batch" (Cotter et al. (2011)) and do several iterations to speed up the gradient descent.
        - Use an adaptive learning rate: high at the beginning and low after, or to take into account the error rate as in (Kuncheva and Plumpton (2008))
    - See results for more specific information.

2. **RGNBC: Rough Gaussian Naïve Bayes Classifier for Data Stream Classification with Recurring Concept Drift**