Multinomial Naive Bayes Classifier

Miguel Alcón Doganoc
Universitat Politècnica de Catalunya
Barcelona, Spain
miguel.alcon@est.fib.upc.edu

1 INTRODUCTION

The Multinomial Naive Bayes classifier (MNB) is part of the family of the Naive Bayes classifiers, which are simple probabilistic classifiers based on applying Bayes's theorem with strong (naive) independence assumptions between the features.

The main characteristic of MNB is that it works with transactional data. This means that each observation of the data is seen as a data structure itself, so it can contain different number of features (items) in each of them.

The goal of MNB is

2 ALGORITHM

Now is time to explain the algorithm behind the MNB. For giving a more practical view, we divided it in two processes, ***fit*** and ***predict***. In the ***fit*** stage, the classifier model is build from the input data, while in the ***predict*** stage, MNBC uses the model to predict an input observation.

2.1 Algorithm of *fit*

The algorithm, described below, makes the necessary counting of the classes c_k and items t_i of the input, in order to compute the probabilities $Pr(c_k)$ and $Pr(t_i \mid c_k)$ of each item and class. In the case of $Pr(t_i \mid c_k)$, we applied the Laplace Correction.

All these probabilities together with the information of the classes form what we call a *model* of the input, which is used in the ***predict*** stage.

Given input X
    Let C be the number of classes in X
    Let c_k be a class of observation,
        with $1 \geq k \geq C$
    Let fc_k be the frequency of class c_k within X
    Let T be the number of unique items in X
    Let t_i be an item of the observations in X,
        with $1 \geq i \geq T$
    Let ft_i be the frequency of item t_i within X
    Let $f_{i,k}$ be the frequency of item t_i
        within observations of class c_k
    For each c_k classes in X
        $Pr(c_k) = \frac{fc_k}{\sum_j^C fc_j}$
        For each t_i items in X
            $Pr(t_i \mid c_k) = \frac{f_{i,k}+1}{\sum_j^T f_{j,k}+T}$

2.2 Algorithm of *predict*

The algorithm of ***predict*** computes $L(c_k)$ for every possible class c_k in order to predict which of them is the class of the observation, with the highest probability (highest $L(c_k)$). This function, $L(c_k)$, comes from

$$Pr(O \mid c_k) = Pr(c_k) \cdot \prod_i^T Pr(t_i \mid c_k)$$

But instead of the multiplication of probabilities, we used the summation of the negative logarithms of the probabilities. This decision was made to avoid the high risk of underflow that the multiplication of small numbers has.

Given an observation O
    Let T be the number of items in O
    Let t_i be an item of O
        with $1 \geq i \geq$ T
    For each c_k possible class
        $L(c_k) = -\log Pr(c_k) + \sum_i^T -\log Pr(t_i \mid c_k)$
    Output the class $argmax_k L(c_k)$

3 DATA

4 IMPLEMENTATION

5 EXPERIMENTATION

6 CONCLUSIONS