ML01 LAB03 ASSIGNMENT

SREELAKSHMI CV
21BDA39

WRITE THE DIFFERENCE BETWEEN THE FOLLOWING

A) Gaussian Naive Bayes

When working with continuous data, an assumption often taken is that the continuous values associated with each class are distributed according to a normal (or Gaussian) distribution. The likelihood of the features is assumed to be-

$$P(x_i \mid y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right)$$

Sometimes assume variance

- is independent of Y (i.e., σi),
- or independent of Xi (i.e., σk)
- or both (i.e., σ)

B) Multinomial Naive Bayes

The Multinomial Naive Bayes algorithm is a Bayesian learning approach popular in Natural Language Processing (NLP). The program guesses the tag of a text, such as an email or a newspaper story, using the Bayes theorem. It calculates each tag's likelihood for a given sample and outputs the tag with the greatest chance

The Naive Bayes method is a strong tool for analyzing text input and solving problems with numerous classes. Because the Naive Bayes theorem is based on the Bayes theorem, it is necessary to first comprehend the Bayes theorem notion. The Bayes theorem, which was developed by Thomas Bayes, estimates the likelihood of occurrence based on prior knowledge of the event's conditions. When predictor B itself is available, we calculate the likelihood of class A. It's based on the formula below: $P(A|B) = P(A) * P(B|A)/P(B)$.

C) Complement Naive Bayes

Complement Naive Bayes is somewhat an adaptation of the standard Multinomial Naive Bayes algorithm. Multinomial Naive Bayes does not perform very well on imbalanced datasets. **Imbalanced datasets** are datasets where the number of examples of some class is higher than the number of examples belonging to other classes. This means that the distribution of examples is not uniform.

In complement Naive Bayes, instead of calculating the probability of an item belonging to a certain class, we calculate the probability of the item belonging to all the classes. This is the literal meaning of the word, complement and hence is called Complement Naive Bayes.

D) Bernoulli Naive Bayes

Bernoulli Naive Bayes, implements the naive Bayes training and classification algorithms for data that is distributed according to multivariate Bernoulli distributions; i.e., there may be multiple features but each one is assumed to be a binary-valued (Bernoulli, boolean) variable. Therefore, this class requires samples to be represented as binary-valued feature vectors; if handed any other kind of data, a `BernoulliNB` instance may binarize its input (depending on the `binarize` parameter).

E) Categorical Naive Bayes,

It is suitable for classification with discrete features which assumes categorically distribution for each feature. The features should to encoded using label encoding techniques such that each category would be mapped to a unique number.

The probability of category `t` in feature `i` given class `c` is estimated as:

$$P(x_i = t \mid y = c \,;\, \alpha) = \frac{N_{tic} + \alpha}{N_c + \alpha n_i},$$

where

Ntic is the number of times category t appears in the samples *xi*, which belong to class *c*,
Nc is the total number of samples with class c,

α is a Laplace smoothing parameter to handle zero frequency problem and

ni is the number of available categories of feature. All unique categories in the data set.

F) Out-of-Core naive Bayes model fitting

Naive Bayes models can be used to tackle large scale classification problems for which the full training set might not fit in memory. To handle this case, **`Multinomial`**, **`BernolliNB`**, and **`Gausiannb`** expose a `partial_fit` method that can be used incrementally as done with other classifiers as demonstrated in out-of-core classification of text documents. All naive Bayes classifiers support sample weighting.

Contrary to the `fit` method, the first call to `partial_fit` needs to be passed the list of all the expected class labels.

What is Jaccard and Cosine Similarity?

Jaccard Similarity is a common proximity measurement used to compute the similarity between two objects, such as two text documents. Jaccard similarity can be used to find the similarity between two asymmetric binary vectors or to find the similarity between two sets. In literature, Jaccard similarity, symbolized by J,can also be referred to as Jaccard Index, Jaccard Coefficient, Jaccard Dissimilarity, and Jaccard Distance.

Jaccard Similarity is frequently used in data science applications. Example use cases for Jaccard Similarity:

- **Text mining: find the similarity between two text documents using the number of terms used in both documents**
- **E-Commerce: from a market database of thousands of customers and millions of items, find similar customers via their purchase history**
- **Recommendation System: Movie recommendation algorithms employ the Jaccard Coefficient to find similar customers if they rented or rated highly many of the same movies.**

Cosine similarity measures the similarity between two vectors of an inner product space. It is measured by the cosine of the angle between two vectors and determines whether two vectors are pointing in roughly the same direction. It is often used to measure document similarity in text analysis.