

IIT Madras
ONLINE DEGREE

Machine Learning Techniques
Professor Doctor Ashish Tendulkar
Indian Institute of Technology Madras
Naive Bayes Classifier

(Refer Slide Time: 00:09)

Naive Bayes Classifier

Namaste! Welcome to the next video of Machine Learning Practice course. This is a short video about Naive Bayes Classifier and how sklearn implements various Naive Bayes classifiers.

(Refer Slide Time: 00:23)

Naive Bayes classifier

- Naive Bayes classifier applies [Bayes' theorem](#) with the “naive” assumption of conditional independence between every pair of features given the value of the class variable.

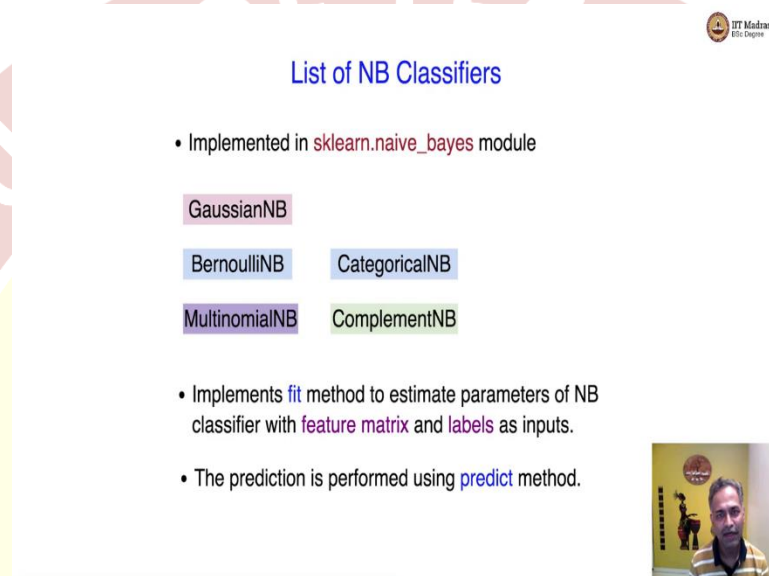
For a given class variable y and dependent feature vector x_1 through x_m , the naive conditional independence assumption is given by:

$$P(x_i|y, x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_m) = P(x_i|y)$$

Naive Bayes learners and classifiers can be extremely fast compared to more sophisticated methods.

We know that Naive Bayes classifier uses Bayes' theorem for predicting the probability of a sample belonging to a specific class. It makes use of a conditional independence assumption among the features. Naive Bayes learners and classifiers can be extremely fast compared to more sophisticated methods.

(Refer Slide Time: 00:50)



List of NB Classifiers

- Implemented in `sklearn.naive_bayes` module

GaussianNB
BernoulliNB CategoricalNB
MultinomialNB ComplementNB

- Implements `fit` method to estimate parameters of NB classifier with `feature matrix` and `labels` as inputs.
- The prediction is performed using `predict` method.

Let us look at list of Naive Bayes classifiers implemented in sklearn. The Naive Bayes classifiers are implemented in `sklearn.naive_bayes` module. There are five different naive bayes classifiers, GaussianNB, BernoulliNB, CategoricalNB, MultinomialNB and ComplementNB that are implemented as part of the Naive Bayes module. So, we use Gaussian Naive Bayes whenever we have continuous features.

If our features are binary then we use Bernoulli Naive Bayes classifier. If our features are distributed according to categorical distribution we use categorical Naive Bayes classifier. In applications like text classification where we assume multinomial distribution multinomialNB is used. And when we have class imbalance issues in the dataset we use complementNB.

So, in case of class imbalance complementNB has been found to be performing better than multinomialNB. Naive Bayes classifiers implement quick method to estimate parameter with feature matrix and labels as input. The prediction is performed using predict method. So, in this short video we studied different Naive Bayes classifier implemented in sklearn.