MIT Art Design and Technology University’s MIT School of Computing, Pune

Department of Computer Science and Engineering

# BTech Third Year A.Y.2023-24

**Artificial Intelligence and Machine Learning Lab**

**Enrollment No:MITU21BTCS0517 Division: TY CORE-1 Roll No:2213041 Name: Samarth Nimbargi**

Experiment No.: 07

**Title:** Develop a Bayesian classifier.

# Theory:

Bayes Theorem

Bayes’ Theorem is a way of finding a probability when we know certain other probabilities.

The formula is:

P(A|B) = *P(A) P(B|A)***P(B)**

|  |  |
| --- | --- |
| Which tells us: | how often A happens *given that B happens*, written **P(A|B)**, |
| When we know: | how often B happens *given that A happens*, written **P(B|A)** |

and how likely A is on its own, written **P(A)**

and how likely B is on its own, written **P(B) Bayes Classifier with example**

In machine learning, naïve Bayes classifiers are a family of simple "probabilistic classifiers" based on applying Bayes' theorem with strong (naïve) independence assumptions between the features.

They are among the simplest Bayesian network models.[1] But they could be coupled with Kernel density estimation and achieve higher accuracy levels.

Naive Bayes is a simple technique for constructing classifiers: models that assign class labels to problem instances, represented as vectors of feature values, where the class labels are drawn from some finite set. There is not a single algorithm for training such classifiers, but a family of algorithms

MIT Art Design and Technology University’s MIT School of Computing, Pune

Department of Computer Science and Engineering

# BTech Third Year A.Y.2023-24

**Artificial Intelligence and Machine Learning Lab**

based on a common principle: all naive Bayes classifiers assume that the value of a

particular feature is independent of the value of any other feature, given the class variable. For example, a fruit may be considered to be an apple if it is red, round, and about 10 cm in diameter. A naive Bayes classifier considers each of these features to contribute independently to the probability that this fruit is an apple, regardless of any possible correlations between the color, roundness, and diameter features.

MIT Art Design and Technology University’s MIT School of Computing, Pune

Department of Computer Science and Engineering

# BTech Third Year A.Y.2023-24

**Artificial Intelligence and Machine Learning Lab**

|  |  |  |
| --- | --- | --- |
|  | **Exercise for Practice** | |
| **Q1.** | | **How can Bayes classifier can be used for categorical features? What if some features are numerical?** |
| Ans: | | **Handling Categorical Features:**   * Count occurrences to calculate likelihood.   **Handling Numerical Features (Gaussian Naive Bayes):**   1. **Calculate Mean and Variance for Each Class:**    * For each class, calculate the mean and variance of each numerical feature. 2. **Gaussian Probability Density Function:**    * Use the Gaussian probability density function to calculate the likelihood of a given feature value for a given class. 3. **Apply Bayes' Theorem:**    * After obtaining the likelihood for each feature, apply Bayes' theorem as usual to calculate the posterior probability. |
| **Q2.** | | **What are advantages and disadvantages of Naive Bayes Algorithm?** |
| Ans: | | **Advantages of Naive Bayes Algorithm:**   1. **Simple and Easy to Implement:**    * Naive Bayes is easy to understand and implement, making it suitable for small datasets and as a baseline algorithm. 2. **Efficient with Large Datasets:**    * It is computationally efficient and can scale well with the size of the dataset. 3. **Works Well with Categorical Features:**    * Particularly effective with categorical features and is robust to irrelevant ones.   **Disadvantages of Naive Bayes Algorithm:**   1. **Assumption of Independence:**    * It assumes that all features are independent, which might not be the case in real-world data. 2. **Sensitive to Feature Distribution:**    * It can be sensitive to the distribution of the data. 3. **Cannot Learn Relationships Between Features:**    * Unable to learn the relationship between features, so the order of features or words is not taken into account. |
| **Q3.** | | **How is Bayes’ Theorem relevant to the field of artificial intelligence?** |

|  |  |
| --- | --- |
| Ans: | **Machine Learning:**   * **Naive Bayes Classifier: Widely used for text classification, spam filtering, and recommendation systems.** * **Bayesian Networks: Utilized for reasoning under uncertainty, e.g., medical diagnosis, risk assessment.** * **Bayesian Optimization: Useful for hyperparameter optimization in machine learning models.**   **Reinforcement Learning:**   * **Bayesian Reinforcement Learning: Used to update beliefs about the environment based on observed outcomes, influencing the decision- making process.**   **Natural Language Processing (NLP):**   * **Text Classification and Sentiment Analysis: Naive Bayes is fundamental for these tasks.**   **Robotics:**   * **Sensor Fusion: Combines data from different sensors to form a more accurate picture of the robot's environment.**   **Medical Diagnosis:**   * **Disease Prediction: Used to calculate the probability of a patient having a disease based on observed symptoms.** |
| **Q4.** | **What are the Applications of Naive Bayes Classifier?** |
| Ans: | The Naive Bayes classifier has various applications across different domains due to its simplicity and effectiveness. Here are some of its key applications:   1. **Text Classification:**    * **Spam Filtering:** Classify emails as spam or non-spam.    * **Sentiment Analysis:** Analyze and classify text sentiment as positive, negative, or neutral.    * **Document Classification:** Categorize documents into predefined categories. 2. **Recommendation Systems:**    * **Product Recommendations:** Recommend products based on user preferences.    * **Content Recommendations:** Suggest articles, videos, or other content based on user behavior. 3. **Medical Diagnosis:**    * **Disease Prediction:** Predict diseases based on symptoms.    * **Patient Health Monitoring:** Analyze patient data to monitor health conditions. 4. **Customer Support:**    * **Ticket Routing:** Route support tickets to the appropriate department.    * **Chatbots:** Classify user queries and respond accordingly. 5. **Fraud Detection:**    * **Credit Card Fraud Detection:** Identify fraudulent credit card transactions.    * **Insurance Claim Fraud Detection:** Detect fraudulent insurance claims. 6. **News Categorization:**    * **Categorize News Articles:** Classify news articles into different categories such as politics, sports, or entertainment. |
| **Q5.** | **Discuss a situation where the naive Bayes’ approach would not be appropriate to use.** |

|  |  |
| --- | --- |
| Ans: | **n image recognition, the Naive Bayes approach would not be appropriate due to:**  **Reasons:**   1. **Dependent Features:**    * **Image pixels are not independent, violating Naive Bayes' assumption of independence.** 2. **High Dimensionality:**    * **High-dimensional image data makes accurate probability estimation challenging.** 3. **Complexity of the Model:**    * **Images have complex structures, which Naive Bayes cannot effectively model.**   **Alternative Approaches:**   * **Convolutional Neural Networks (CNNs):**   + **Designed for image data, CNNs can effectively learn and extract features from images.** * **Deep Learning Models:**   + **Deep learning models are well-suited for image recognition tasks due to their ability to learn complex patterns and relationships directly from raw data.** * **SVMs with Nonlinear Kernels:**   + **SVMs with nonlinear kernels can be an alternative for image recognition tasks.** * **Decision Trees or Random Forests:**   + **These methods can also be used for image recognition, capable of capturing non-linear relationships.** |
| **Conclusion** | Thus, we have successfully completed the implementation of Naïve BayesGaussian Classifier. |

