1.Given the following data about weather and playing tennis:

Outlook: Sunny, Temperature: Hot, Humidity: High, Windy: False, Play: No

Outlook: Rain, Temperature: Mild, Humidity: Normal, Windy: False, Play: Yes

Calculate the probability of playing tennis using Naive Bayes classifier. Show all steps.

To calculate the probability of playing tennis using the Naive Bayes classifier, we need to compute the conditional probabilities based on the given data.

Given Data:

- **Sunny, Hot, High, False**: Play = No
- **Rain, Mild, Normal, False**: Play = Yes

Step 1: Calculate Prior Probabilities

- Total examples = 2
- P(Play = Yes) = Number of Yes / Total = $\frac{1}{2}$
- P(Play = No) = Number of No / Total = $\frac{1}{2}$

Step 2: Calculate Likelihoods We need to calculate the likelihood of each feature given the class.

- For **Play = Yes**:
  - P(Outlook = Rain | Play = Yes) = 1 (since all examples with Play = Yes have Outlook = Rain)
  - P(Temperature = Mild | Play = Yes) = 1
  - P(Humidity = Normal | Play = Yes) = 1
  - P(Windy = False | Play = Yes) = 1
- For **Play = No**:
  - P(Outlook = Sunny | Play = No) = 1
  - P(Temperature = Hot | Play = No) = 1
  - P(Humidity = High | Play = No) = 1
  - P(Windy = False | Play = No) = 1

Step 3: Calculate Posterior Probability To find the probability of playing tennis given the features, we use Bayes' theorem:

$ P(Play = Yes | \text{features}) = \frac{P(\text{features} | Play = Yes) \cdot P(Play = Yes)}{P(\text{features})} $

Since we only have two classes, we can compare:

$ P(Play = Yes | \text{features}) \propto P(\text{features} | Play = Yes) \cdot P(Play = Yes) $ $ P(Play = No | \text{features}) \propto P(\text{features} | Play = No) \cdot P(Play = No) $

Calculating for both classes:

- For **Play = Yes**: $P(\textbf{features}|Play = Yes) = 1 \cdot 1 \cdot 1 \cdot 1 = 1$
  $P(Play = Yes|\textbf{features}) \propto 1 \cdot \frac{1}{2} = \frac{1}{2}$
- For **Play = No**: $P(\textbf{features}|Play = No) = 1 \cdot 1 \cdot 1 \cdot 1 = 1$
  $P(Play = No|\textbf{features}) \propto 1 \cdot \frac{1}{2} = \frac{1}{2}$

Since both probabilities are equal, we cannot determine a preference based on this

2.Derive the EM algorithm steps for a mixture of two Gaussian distributions.

Explain how it handles missing data.

2. EM Algorithm Steps for Mixture of Two Gaussian Distributions

The Expectation-Maximization (EM) algorithm consists of two main steps: the Expectation step (E-step) and the Maximization step (M-step).

E-step:

- Calculate the expected value of the log-likelihood function, with respect to the current estimate of the distribution parameters. For a mixture of two Gaussians, this involves calculating the responsibilities (probabilities) that each Gaussian component is responsible for each data point.

M-step:

- Update the parameters of the Gaussian distributions (means, variances, and mixing coefficients) based on the responsibilities calculated in the E-step.

Handling Missing Data: The EM algorithm can handle missing data by treating the missing values as latent variables. During the E-step, the algorithm estimates the missing data based on the current parameters, and in the M-step, it updates the parameters using both the observed and estimated missing data.

3. Calculate the sample complexity for learning a threshold function on the real line with error ε and confidence δ. Show your working.

3. Sample Complexity for Learning a Threshold Function

To determine the sample complexity for learning a threshold function on the real line with error $ \epsilon $ and confidence $ \delta $, we can use the following reasoning:

Definition of Sample Complexity

Sample complexity refers to the number of training examples required to ensure that the learning algorithm can achieve a certain level of accuracy (error) with a specified level of confidence.

Threshold Function

A threshold function can be defined as follows:

- It outputs one class (e.g., YES) if the input is greater than a certain threshold t and another class (e.g., NO) if it is less than or equal to t.

Error and Confidence

- **Error ϵ**: This is the maximum allowable probability of making an incorrect prediction.
- **Confidence δ**: This is the probability that the learning algorithm will perform within the error bound ϵ.

Sample Complexity Formula

The sample complexity $ N $ can be derived from the VC (Vapnik-Chervonenkis) dimension of the hypothesis class. For threshold functions on the real line, the VC dimension is 1. The formula for the sample complexity is given by:
$ N \geq \frac{1}{\epsilon} \left(\log \frac{1}{\delta} + \log \frac{1}{\epsilon} \right) $

Explanation of the Formula

- The term $\frac{1}{\epsilon}$ indicates that as the allowable error decreases, the number of samples required increases.
- The term $\log \frac{1}{\delta}$ accounts for the confidence level; as we want higher confidence, we need more samples.
- The term $\log \frac{1}{\epsilon}$ reflects the relationship between the error and the number of samples needed to ensure that the error remains within the specified bounds.

4.Design a Bayesian network for a student's performance prediction system considering factors like study hours, previous grades, attendance, and difficulty level. Calculate conditional probabilities for a given scenario.

3. Sample Complexity for Learning a Threshold Function

To determine the sample complexity for learning a threshold function on the real line with error $ \epsilon $ and confidence $ \delta $, we can use the following reasoning:

Definition of Sample Complexity

Sample complexity refers to the number of training examples required to ensure that the learning algorithm can achieve a certain level of accuracy (error) with a specified level of confidence.

Threshold Function

A threshold function can be defined as follows:

- It outputs one class (e.g., YES) if the input is greater than a certain threshold t and another class (e.g., NO) if it is less than or equal to t.

Error and Confidence

- **Error ϵ**: This is the maximum allowable probability of making an incorrect prediction.
- **Confidence δ**: This is the probability that the learning algorithm will perform within the error bound ϵ.

Sample Complexity Formula

The sample complexity $ N $ can be derived from the VC (Vapnik-Chervonenkis) dimension of the hypothesis class. For threshold functions on the real line, the VC dimension is 1. The formula for the sample complexity is given by:
$ N \geq \frac{1}{\epsilon} \left(\log \frac{1}{\delta} + \log \frac{1}{\epsilon} \right) $

Explanation of the Formula

- The term $\frac{1}{\epsilon}$ indicates that as the allowable error decreases, the number of samples required increases.
- The term $\log \frac{1}{\delta}$ accounts for the confidence level; as we want higher confidence, we need more samples.
- The term $\log \frac{1}{\epsilon}$ reflects the relationship between the error and the number of samples needed to ensure that the error remains within the specified bounds.

Conditional Probability Tables (CPTs)

Each node in the Bayesian network has an associated conditional probability table that quantifies the effect of its parent nodes on its probability distribution. For example:

- $P(P|S,G,A,D)$: The probability of passing based on study hours, previous grades, attendance, and difficulty level.

Example Scenario

Suppose we want to calculate the probability of a student passing the course given the following conditions:

- Study Hours = 5
- Previous Grades = B
- Attendance = 90%
- Difficulty Level = Medium

To compute $ P(P = Pass | S = 5, G = B, A = 90%, D = Medium) $, we would use the conditional probabilities from the CPTs associated with each node. The calculation would involve multiplying the relevant probabilities based on the structure of the network.

Conclusion

The Bayesian network provides a structured way to model the dependencies between various factors affecting a student's performance, allowing for probabilistic inference based on observed data.

5.If S is a collection of 14 examples with 9 YES and 5 NO examples in which one of the attributes is wind speed. The values of Wind can be Weak or Strong. The classification of these 14 examples are 9 YES and 5 NO. For attribute Wind, suppose there are 8 occurrences of Wind = Weak and 6 occurrences of Wind = Strong. For Wind = Weak, 6 of the examples are YES and 2 are NO. For Wind = Strong, 3 are YES and 3 are NO. Find the Entropy(weak) and Entropy(strong).

5. Entropy Calculation for Wind Attribute

To calculate the entropy for the Wind attribute, we use the formula for entropy:
$ H(X) = -\sum_{i} P(x_i) \log_2 P(x_i) $
For Wind = Weak:

- YES = 6, NO = 2
- Total = 8

$ P(YES | Weak) = \frac{6}{8}, \quad P(NO | Weak) = \frac{2}{8} $
$ H(Weak) = -\left(\frac{6}{8} \log_2 \frac{6}{8} + \frac{2}{8} \log_2 \frac{2}{8} \right) $
Calculating:
$ H(Weak) = -\left(\frac{3}{4} \log_2 \frac{3}{4} + \frac{1}{4} \log_2 \frac{1}{4} \right) $
$ = -\left(\frac{3}{4} \cdot (-0.415) + \frac{1}{4} \cdot (-2) \right) \approx 0.811 $
For Wind = Strong:

- YES = 3, NO = 3
- Total = 6

$ P(YES | Strong) = \frac{3}{6}, \quad P(NO | Strong) = \frac{3}{6} $
$ H(Strong) = -\left(\frac{3}{6} \log_2 \frac{3}{6} + \frac{3}{6} \log_2 \frac{3}{6} \right) $
Calculating:
$ H(Strong) = -2 \cdot \left(\frac{1}{2} \log_2 \frac{1}{2} \right) = 1 $

Summary of Entropy Values

- $\boldsymbol{H(Weak) \approx 0.811}$
- $\boldsymbol{H(Strong) = 1}$

These calculations provide insights into the uncertainty associated with the Wind attribute in predicting whether a student will play tennis.