BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI – HYDERABAD CAMPUS
BITS F464 : MACHINE LEARNING
FIRST SEMESTER 2023-2024

A worked-out example on Naïve Bayes

Here's a step-by-step solved numerical problem on the Naïve Bayes Classifier.

Let's take the dataset of fruits classified as either "Apple" or "Orange" based on two features: "Color" and "Texture":

FRUIT	COLOR	TEXTURE
Apple	Red	Smooth
Apple	Orange	Smooth
Orange	Orange	Bumpy
Orange	Red	Bumpy
Apple	Red	Smooth
Orange	Orange	Smooth
Apple	Red	Bumpy

We want to use the Naive Bayes classifier to predict whether a fruit is an "Apple" or an "Orange" based on the given features "Color" and "Texture".

Step 1: Calculate Class Probabilities $P(Y)$:

Calculate the probability of each class based on the training data.

Let N_{Apple} be the number of Apple fruits (4) and N_{Orange} be the number of Orange fruits (3).

For Apple (Y = "Apple"):

$$P(Y = "Apple") = \frac{N_{Apple}}{N_{Total}} = \frac{4}{7}$$

For Orange (Y = "Orange"):

$$P(Y = "Orange") = \frac{N_{Orange}}{N_{Total}} = \frac{3}{7}$$

Step 2: Calculate Feature Probabilities $P(X_j|Y)$:

Calculate the probability of each feature given each class.

For Feature **"Color"**:

- Given Apple (Y = "Apple"):

$$P(\text{"Red"} \mid Y = \text{"Apple"}) = \frac{Count\ of\ Red\ Apples}{Count\ of\ Apple\ fruits} = \frac{3}{4}$$

$$P(\text{"Orange"} \mid Y = \text{"Apple"}) = \frac{Count\ of\ Orange\ Apples}{Count\ of\ Apple\ fruits} = \frac{1}{4}$$

- Given Orange (Y = "Orange"):

$$P(\text{"Red"} \mid Y = \text{"Orange"}) = \frac{Count\ of\ Red\ Oranges}{Count\ of\ Orange\ fruits} = \frac{1}{3}$$

$$P(\text{"Orange"} \mid Y = \text{"Orange"}) = \frac{Count\ of\ Orange\ Oranges}{Count\ of\ Orange\ fruits} = \frac{2}{3}$$

For Feature **"Texture"**:

- Given Apple (Y = "Apple"):

$$P(\text{"Smooth"} \mid Y = \text{"Apple"}) = \frac{Count\ of\ Smooth\ Apples}{Count\ of\ Apple\ fruits} = \frac{3}{4}$$

$$P(\text{"Bumpy"} \mid Y = \text{"Apple"}) = \frac{Count\ of\ Bumpy\ Apples}{Count\ of\ Apple\ fruits} = \frac{1}{4}$$

- Given Orange (Y = "Orange"):

$$P(\text{"Smooth"} \mid Y = \text{"Orange"}) = \frac{Count\ of\ Smooth\ Oranges}{Count\ of\ Orange\ fruits} = \frac{1}{3}$$

$$P(\text{"Bumpy"} \mid Y = \text{"Orange"}) = \frac{Count\ of\ Bumpy\ Oranges}{Count\ of\ Orange\ fruits} = \frac{2}{3}$$

Step 3: Calculate Predicted Probabilities *P(Y|X)*:

Let us predict whether a fruit *X* with *Color="Orange"* and *Texture="Smooth"* is an "Apple" or an "Orange". Calculate the predicted probabilities for each class given the features using the Naive Bayes formula:

$$P(Y \mid X) = P(Y) \times P(X_1 \mid Y) \times P(X_2 \mid Y)$$

For Apple (Y = "Apple"):

$$\begin{aligned} P(Y = \text{"Apple"} \mid X) &= P(Y = \text{"Apple"}) \times P(\text{"Orange"} \mid Y = \text{"Apple"}) \times P(\text{"Smooth"} \mid Y = \text{"Apple"}) \\ &= \frac{4}{7} \times \frac{1}{4} \times \frac{3}{4} \\ &= 0.1071 \end{aligned}$$

For Orange (Y = “Orange”):

$$P(Y = "Orange"|X) = P(Y = "Orange") \times P("Orange"|Y = "Orange") \times P("Smooth"|Y = "Orange")$$
$$= \frac{3}{7} \times \frac{2}{3} \times \frac{1}{3}$$
$$= 0.0952$$

Step 4: Make Prediction:

Comparing the predicted probabilities, we find that *P(Y = “Apple” | X)* is higher than *P(Y = “Orange” | X)*.

Therefore, we predict that the given fruit (Orange color, Smooth texture) is more likely to be an “Apple”.

Final Prediction: The given fruit is predicted to be an “Apple”.

This concludes the Naive Bayes classification process for the given dataset and test fruit. We calculate the probabilities for each class and make a prediction based on the higher probability.