NAIVE BAYES CLASSIFIER

TEAM MEMBERS

BIVASHA BASHIR BIVA

DHARMIK LOKESH SHAH

MOHAMED RAFI IBRAHIM

Instructor: Dr. Yasser Alginahi

Date: 20th October 2023

University of Windsor

What's in it for you

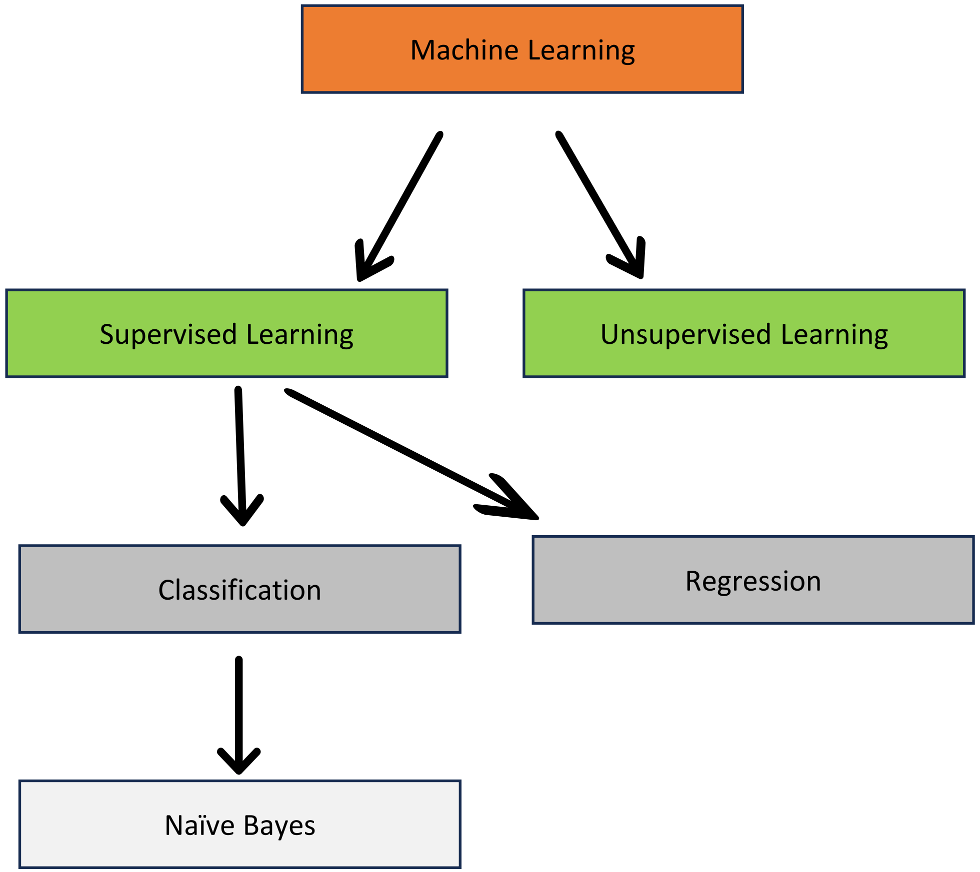
1. Methods to establish Classifier
2. What is Naïve Bayes?
3. Data Preprocessing
4. Introduction to Conditional Probability & Bayes Theorem
5. Zero Frequency Problem
6. Naïve Bayes Algorithm
7. Types of Naïve Bayes Classifier
8. Demo / Example
9. Advantages | Disadvantages of Naive Bayes Classifier
10. Application of Naïve Bayes Classifier
11. References

Methods to establish a classifier

- There are three methods to establish a classifier
  - a) Model a classification rule directly
    - Examples: k-NN, decision trees, perceptron, SVM
  - b) Model the probability of class memberships given input data
    - Example: multi-layered perceptron with the cross-entropy cost
  - c) Make a probabilistic model of data within each class
    - Examples: Naive Bayes, model-based classifiers
- a) and b) are examples of discriminative classification
- c) is an example of generative classification
- b) and c) are both examples of probabilistic classification

INTRODUCTION TO NAÏVE BAYES CLASSIFIER

- Naïve bayes classifier assumes that all the features are unrelated to each other. Presence or absence of a feature does not influence the presence or absence of any other feature.
- It is a Statistical method for classification.
- It Can solve problems involving both categorical and continuous-valued attributes.
- Comes under Supervised Learning.
- Assumes an underlying probabilistic model, the Bayes theorem.



Introducing Naïve Bayes Classifier

- ❖ Principles of conditional probability given by Bayes' theorem
- ❖ Let's understand the simple concepts in probability that are used in this algorithm

DATA PREPROCESSING

If a dataset is non-continuous (categorical or discrete) data, we will need to preprocess it before applying Gaussian Naive Bayes.

Here's how you can handle non-continuous data:

- **Convert Categorical Data**: For categorical or discrete features, you will need to convert them into a continuous representation. This can be done using techniques like one-hot encoding or label encoding.
- **Handle Missing Values**: This can involve imputation with mean, median, or mode for continuous features and using a specific category for missing values in categorical features.

CONDITIONAL PROBABILITY

- In probability theory, conditional probability is a measure of the probability of an event given that another event has already occurred.
- If the event of interest is A and the event B is assumed to have occurred, "the conditional probability of A given B", or "the probability of A under the condition B", is usually written as P(A|B), or sometimes PB(A).

Example 1: Chances of cough

- The probability that any given person has a cough on any given day maybe only 5%. But if we know or assume that the person has a cold, then they are much more likely to be coughing. The conditional probability of coughing given that person have a cold might be a much higher 75%.

Example 2: Probability of getting a blue marble

- 2 blue and 3 red marbles are in a bag. What are the chances of getting a blue marble?

Answer : - The chance is 2/5

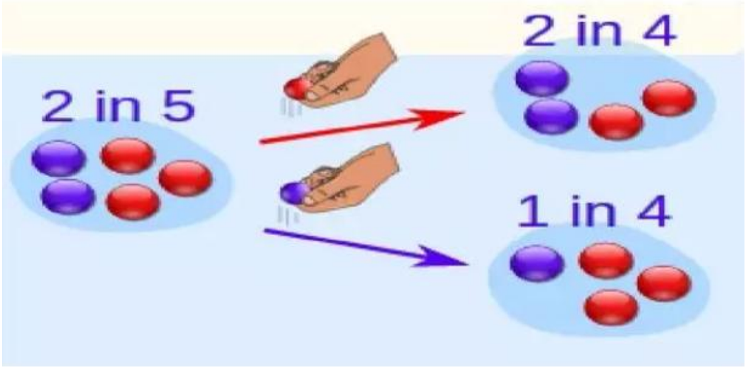


Figure: Probability of getting a blue marble **[3]**

- But after taking one out of these chances, situation , may change!
- So, the next time:
- If we got a red marble before, then the chance of a blue marble next is 2/4
- If we got a blue marble before, then the chance of a blue marble next is 1/4

University of Windsor

Zero Frequency Problem!

What if any of the count is 0?

- Add 1 to all counts
- It is a form of Laplace Smoothing
- See https://en.wikipedia.org/wiki/Additive_smoothing
- **Formula:** The probability of a feature given a class after Laplace Smoothing is calculated as follows:
- P(feature | class) = (Count(feature in class) + 1) / (Total count of features in class + Number of unique features)

University of Windsor

Introducing Bayes Theorem

Bayes' theorem gives the conditional probability of an event A given another event B occurred

Bayes Theorem

$$P(A \mid B) = \frac{P(B \mid A) \cdot P(A)}{P(B)}$$

Where,

- P(A|B) = conditional probability A given B (posterior probability)
- P(B|A) = conditional probability of B given A (likelihood)
- P(A) = probability of event A (class prior probability)
- P(B) = probability of event B (predictor prior probability)

The conditional probability of occurrence of an event based on prior knowledge of conditions that might be related to the event

Types of Naïve Bayes Classifier

- Multinomial Naïve Bayes
- Gaussian Naïve Bayes
- Bernoulli Naïve Bayes

Multinomial Bayesian model:

- A probabilistic classification model.
- Used in text classification.
- Assigns documents to predefined categories.
- Calculates class probability based on word/feature frequency and Bayes' theorem.
- For instance, when classifying movie reviews as "positive" or "negative," feature frequencies involve word counts like "good," "bad," "amazing," and "terrible." The Bayesian model uses these frequencies to determine whether the review falls into the "positive" or "negative" category.

University of Windsor

Example for Multinomial Naïve Bayes:

- Frequency table of the training data set given count of all the Cricket Playing conditions against the respective weather condition

WEATHER	YES	NO
SUNNY	3	2
OVERCAST	4	0
RAINY	2	3
TOTAL	9	5

- Find the probabilities of each weather condition and create a likelihood table

  P(Yes) = 9/14(0.64)
  P(No) = 5/14(0.36)

WEATHER	YES	NO	
SUNNY	3	2	5/14(0.36)
OVERCAST	4	0	4/14(0.29)
RAINY	2	3	5/14(0.36)
TOTAL	9	5	

- After replacing variables in the above formula, we get:
- P(Yes | Sunny) = P(Yes) * P(Sunny | Yes) / P(Sunny)
- Take the values from the above likelihood table and put it in the above formula.
- P(Sunny | Yes) = 3/9 = 0.33, P(Yes) = 0.64 and P(Sunny) = 0.36
- Hence, P(Yes | Sunny) = (0.64*0.33)/0.36 = 0.60
- P(No | Sunny) = P(No) * P(Sunny | No) / P(Sunny)
- Take the values from the above likelihood table and put it in the above formula.
- P(Sunny | No) = 2/5 = 0.40, P(No) = 0.36 and P(Sunny) = 0.36
- P(No | Sunny) = (0.36*0.40)/0.36 = 0.6 = 0.40
- The probability of playing cricket in sunny weather conditions is higher. Hence, the player will play if the weather is sunny.

Gaussian Naïve Bayes

- Assumes that continuous features in dataset follow a Gaussian(Normal) distribution

$$P(X|Y=c) = \frac{1}{\sqrt{2\pi\sigma_c^2}} e^{\frac{-(x-\mu_c)^2}{2\sigma_c^2}}$$

- Variance (σ^2) and Mu (μ) of the continuous variable X computed for a given class c of Y.

Example for Gaussian Naïve Bayes:-

- Let's consider a simple example of a Gaussian Naive Bayes classifier using mathematical terms and calculations:
- Suppose we have a dataset with two features, $X1$ and $X2$, and a binary target variable Y indicating two classes (0 or 1). We want to predict the class label based on these features.
- Let's say we have the following data points:

  Data points: X1:[1.2,0.8,1.0,1.5,1.3]

  X2:[1.0,1.5,1.2,0.9,1.3]

  Y:[1,0,1,0,1]

- To use Gaussian Naive Bayes, we need to calculate the mean (μ) and standard deviation (σ) for each feature in each class. Let's compute these values for both classes (0 and 1).

- For Class 0 (Y=0):
- Mean and Standard Deviation for X1:

$$\mu_{X_1,0} = \frac{1.5+0.9}{2} = 1.2$$
$$\sigma_{X_1,0} = \sqrt{\frac{(1.5-1.2)^2+(0.9-1.2)^2}{2}} \approx 0.2121$$

- Mean and Standard Deviation for X2:

$$\mu_{X_2,0} = \frac{1.5+1.3}{2} = 1.4$$
$$\sigma_{X_2,0} = \sqrt{\frac{(1.5-1.4)^2+(1.3-1.4)^2}{2}} \approx 0.0483$$

- For Class 1 (Y=1):
- Mean and Standard Deviation for X1:

$$\mu_{X_1,1} = \frac{1.2+1.0+1.3}{3} \approx 1.1667$$
$$\sigma_{X_1,1} = \sqrt{\frac{(1.2-1.1667)^2+(1.0-1.1667)^2+(1.3-1.1667)^2}{3}} \approx 0.0527$$

- Mean and Standard Deviation for X2:

$$\mu_{X_2,1} = \frac{1.0+1.2+1.3}{3} \approx 1.1667$$
$$\sigma_{X_2,1} = \sqrt{\frac{(1.0-1.1667)^2+(1.2-1.1667)^2+(1.3-1.1667)^2}{3}} \approx 0.1019$$

- Now, given a new data point (x1,x2)=(1.1,1.4), we can calculate the likelihoods for both classes using the Gaussian probability density function:
- For Class 0 (Y=0):

$$P(X_1 = 1.1|Y = 0) = \frac{1}{\sqrt{2\pi\sigma^2_{X_1,0}}} e^{-\frac{(1.1-\mu_{X_1,0})^2}{2\sigma^2_{X_1,0}}}$$
$$P(X_2 = 1.4|Y = 0) = \frac{1}{\sqrt{2\pi\sigma^2_{X_2,0}}} e^{-\frac{(1.4-\mu_{X_2,0})^2}{2\sigma^2_{X_2,0}}}$$

- For Class 1 (Y=1):

$$P(X_1 = 1.1|Y = 1) = \frac{1}{\sqrt{2\pi\sigma^2_{X_1,1}}} e^{-\frac{(1.1-\mu_{X_1,1})^2}{2\sigma^2_{X_1,1}}}$$

$$P(X_2 = 1.4|Y = 1) = \frac{1}{\sqrt{2\pi\sigma^2_{X_2,1}}} e^{-\frac{(1.4-\mu_{X_2,1})^2}{2\sigma^2_{X_2,1}}}$$

- Finally, we can calculate the posterior probabilities using Bayes' theorem and predict the class with the higher probability as the output class for the given data point.

Bernoulli Naïve Bayes

- Efficient for binary feature tasks like text classification.
- Thrives with limited resources and small datasets.
- Performance may suffer if feature independence is violated or feature frequency matters.
- For instance, while classifying emails into two classes: "spam" and "not spam." You've trained your model on a dataset of emails, and one of the features you're using is the presence or absence of the word "offer." The Bayesian model uses this features to determine whether the Email is a "Spam" or "Ham".

Example for Bernoulli Naïve Bayes :-

- In the Below dataset, we are trying to predict whether a person has a disease or not based on their age, gender, and fever. Here, 'Disease' is the target, and the rest are the features.

Adult	Gender	Fever	Disease
Yes	Female	No	False
Yes	Female	Yes	True
No	Male	Yes	False
No	Male	No	True
Yes	Male	Yes	True

Figure: Diagnosis of fever and Disease [5]

- All values are binary. We wish to classify an instance 'X' where Adult='Yes', Gender= 'Male', and Fever='Yes'.

- Firstly, we calculate the class probability, probability of disease or not.
  P(Disease = True) = ⅗
  P(Disease = False) = ⅖
- Secondly, we calculate the individual probabilities for each feature.
  P(Adult= Yes | Disease = True) = ⅔
  P(Gender= Male | Disease = True) = ⅔
  P(Fever= Yes | Disease = True) = ⅔

  P(Adult= Yes | Disease = False) = ½
  P(Gender= Male | Disease = False) = ½
  P(Fever = Yes | Disease = False) = ½

- Now, we need to find out two probabilities:-

  (i) $P(Disease{=}\ True \mid X) = (P(X \mid Disease{=}\ True) * P(Disease{=}True))/\ P(X)$

  (ii) $P(\ Disease = False \mid X) = (P(X \mid Disease = False) * P(Disease{=}\ False)\ )/P(X)$

  $P(Disease = True \mid X) = ((\ \frac{2}{3} * \frac{2}{3} * \frac{2}{3}\ ) * (\frac{3}{5}))/P(X) = (8/27 * \frac{3}{5}) / P(X)\ = 0.17/P(X)$

  $P(Disease = False \mid X) = [(\frac{1}{2} * \frac{1}{2} * \frac{1}{2}\ ) * (\frac{2}{5})] / P(X)\ = [\frac{1}{8} * \frac{2}{5}] / P(X)\ = 0.05/\ P(X)$

- Now, we calculate estimator probability:-

  $P(X) = P(Adult{=}\ Yes) * P(Gender = Male\ ) * P(Fever = Yes)$

  $= \frac{3}{5} * \frac{3}{5} * \frac{3}{5} = 27/125 = 0.21$

- So we get finally,

  $P(Disease = True \mid X) = 0.17 / P(X)$

  $= 0.17 / 0.21$

  $= 0.80$ -------- (eq-1)

- $P(Disease = False \mid X) = 0.05 / P(X)$

  $= 0.05 / 0.21$

  $= 0.23$ -------- (eq-2)

- Now, we notice that (1) > (2), the result of instance 'X' is 'True', i.e., the person has the disease.

How Naïve Bayes algorithm works (Example)

- We are training dataset of weather and corresponding target variable 'Play'.
- Now, we need to classify whether players will play the cricket match or not based on weather condition.

Weather	Play (YES / NO)
Sunny	NO
Overcast	YES
Rainy	YES
Sunny	YES
Sunny	YES
Overcast	YES
Rainy	NO
Rainy	NO
Sunny	YES
Rainy	YES
Sunny	NO
Overcast	YES
Overcast	YES
Rainy	NO

Figure: play condition based on weather. [6]

1. Convert the dataset into a frequency table.

Weather	Play
Sunny	No
Overcast	Yes
Rainy	Yes
Sunny	Yes
Sunny	Yes
Overcast	Yes
Rainy	No
Rainy	No
Sunny	Yes
Rainy	Yes
Sunny	No
Overcast	Yes
Overcast	Yes
Rainy	No

Frequency Table		
Weather	No	Yes
Overcast		4
Rainy	3	2
Sunny	2	3
Grand Total	5	9

Figure: play condition based on weather. (6)

2. Create likelihood table by finding the probabilities like Overcast probability = 0.29 and probability of playing is 0.64.
3. Now, use the Naïve Bayesian equation to calculate the posterior probability for each class. The class with the highest posterior probability is the outcome of prediction.

Likelihood table				
Weather	No	Yes		
Overcast		4	=4/14	0.29
Rainy	3	2	=5/14	0.36
Sunny	2	3	=5/14	0.36
All	5	9		
	=5/14	=9/14		
	0.36	0.64		

Figure: play condition based on weather.[6]

Problem: Players will play if weather is sunny. Is this statement correct?

P(Yes | Sunny) = P(Sunny | Yes)*P(Yes)/P(Sunny)

Here we have,

P(Sunny | Yes) = 3/9 = 0.33, P(Sunny) = 5/14 = 0.36,

P(Yes) = 9/14 = 0.64

Now, P(Yes | Sunny) = 0.33 * 0.64 / 0.36 = 0.60, (high probability)

Problem: Players will play if weather is sunny. Is this statement correct?

P(No | Sunny) = P(Sunny | No)*P(No)/P(Sunny)

Here we have,

P(Sunny | No) = 2/5 = 0.4,   P(Sunny) = 5/14 = 0.36,

P(No) = 5/14 = 0.36

Now, P(No | Sunny) = 0.4 * 0.36 / 0.36 = 0.40, (low probability)

DEMO | A sample code describing Gaussian NB

- Standard Libraries like numpy, pandas and matplotlib are used to perform numeric operations, managing data and visualizing the analysis of data

```python
# @title Default title text
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

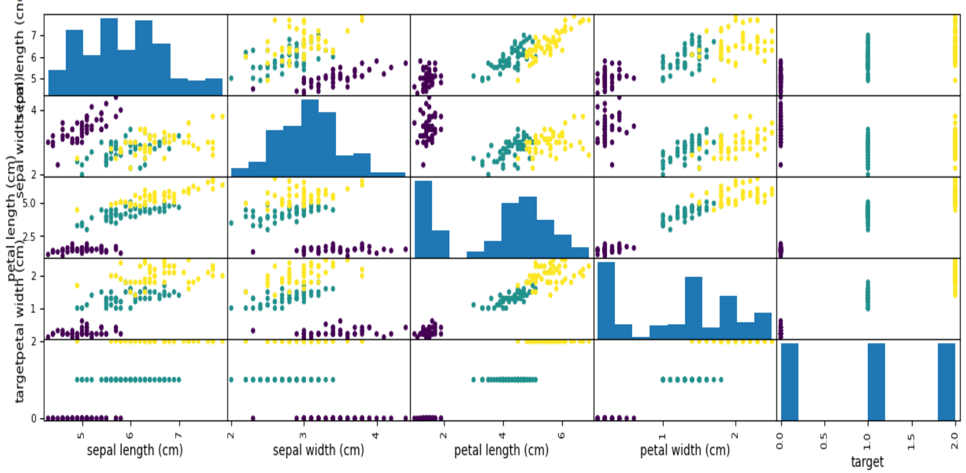
- Loading the Dataset and Visualizing the scatter-plot features

```python
from sklearn import datasets
iris = datasets.load_iris()

X = iris.data[:,]
y=iris.target

print ("Features : ", iris['feature_names'])
iris_dataframe = pd.DataFrame(data=np.c_[iris['data'],iris['target']], columns=iris['feature_names']+['target'])
plt.figure()
grr = pd.plotting.scatter_matrix(iris_dataframe, c=iris['target'], figsize=(15,5), s=60, alpha=1)
plt.show()
```

```
Features :  ['sepal length (cm)', 'sepal width (cm)', 'petal length (cm)', 'petal width (cm)']
<Figure size 640x480 with 0 Axes>
```

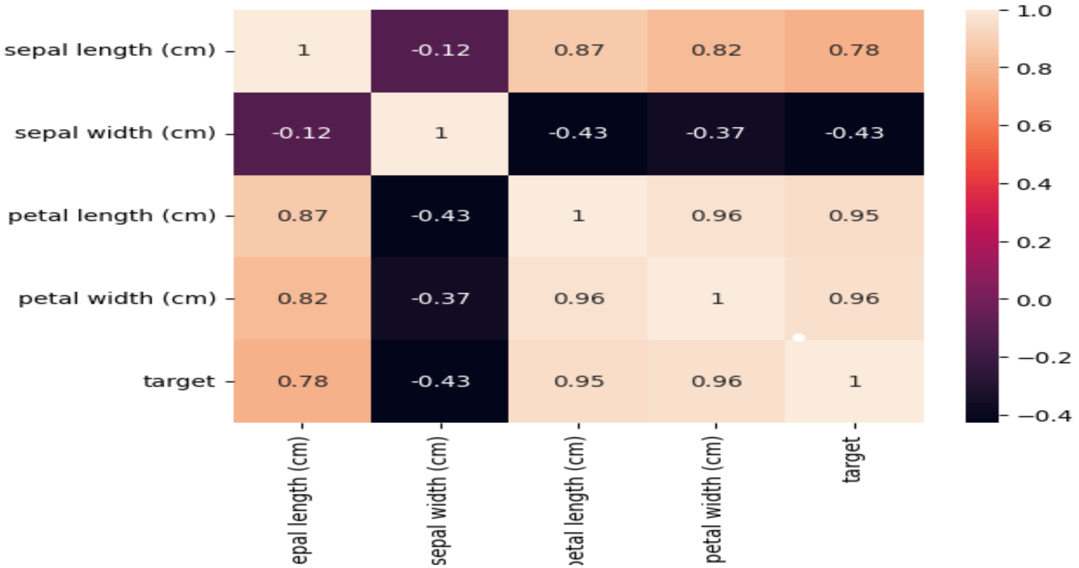


A sample code describing Gaussian NB

- Visualizing the correlation and checking the assumption of Naïve Bayes

```python
import seaborn as sns
dataplot = sns.heatmap(iris_dataframe.corr(), annot=True)
plt.show()
```

- From this we can see the features are highly correlated. But as per Naïve Bayes assumption, it will treat features as independent of each other. Based on this, our algorithm will compute the following probability of all 3 flower classes



$$p(setosa \mid SL, SW, PL, PW) \propto p(SL \mid setosa) * p(SW \mid setosa) * p(PL \mid setosa) * p(PW \mid setosa) * p(setosa)$$

A sample code describing Gaussian NB

- Train the model

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.25, random_state=42)
```

```
[ ]
from sklearn.naive_bayes import GaussianNB

NB = GaussianNB()
NB.fit(X_train, y_train)
```

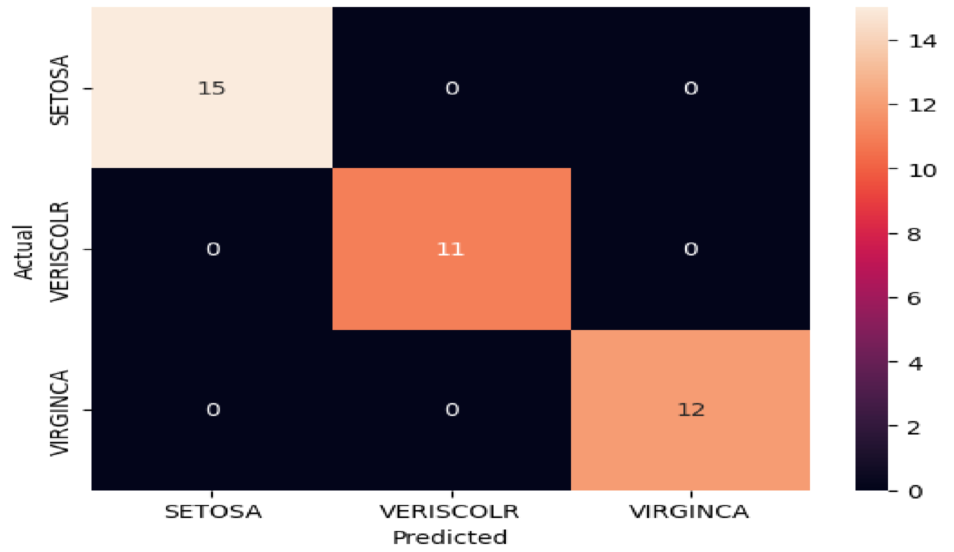
Predict the Model Output

```
y_pred = NB.predict(X_test)

from sklearn.metrics import confusion_matrix
cm = confusion_matrix(y_test, y_pred)
df_cm = pd.DataFrame(cm, columns=['SETOSA', 'VERISCOLR', 'VIRGINCA'],
                     index= ['SETOSA', 'VERISCOLR', 'VIRGINCA'])

df_cm.index.name= 'Actual'
df_cm.columns.name = 'Predicted'

sns.heatmap(df_cm, annot=True )
plt.show()
```



University of Windsor

A sample code describing Gaussian NB

The confusion Matrix allows us to measure Recall and Precision, which, along with Accuracy and the AUC-ROC curve, are the metrics used to measure the performance of ML models.

- **Precision** (positive predicted) is the fraction of relevant data among the retrieved data. On the other hand, **recall** (sensitivity) is the fraction of relevant data that were retrieved. Both precision and recall are based on relevance.
- F1 score is more useful than accuracy, especially when you have an uneven class distribution. The F1-score captures both the trends in a single value:

$$F1 - score = \frac{2}{\frac{1}{Recall} + \frac{1}{Precision}}$$

```
========== Naive Bayes Results ==========
Accuracy     :  1.0
Recall       :  1.0
Precision    :  1.0
F1 Score     :  1.0
Confusion Matrix:
[[10  0  0]
 [ 0 10  0]
 [ 0  0 10]]
```

YOUTUBE VIDEO

- For understanding Gaussian Naïve Bayes Algorithm

https://youtu.be/l3dZ6ZNFjo0?si=e5mxd0EJv7XT_Om6

- To understand Bayesian theorem and how its used in Naïve Bayes

https://youtu.be/lFJbZ6LVxN8?si=RTp07Dp1G4_g-wOc

- A detailed code description from scratch to build Naïve Bayes Classifier https://tinyurl.com/45wk35u8

ADVANTAGES

- ❖Very simple and easy to implement
- ❖Needs less training data
- ❖Handles both continuous and discrete data
- ❖Highly scalable with number of predictors and data science
- ❖Not sensitive to irrelevant features

DISADVANTAGES

- ❖ They are quick and easy to implement but their biggest disadvantage is the requirement for predictors needs to be independent.

APPLICATIONS

Naive Bayes can work surprisingly well in many practical applications, especially when you have limited data or want a simple and interpretable model.

- **Document Categorization:** Naive Bayes is used for categorizing documents into predefined categories or topics, such as news articles, research papers, or user reviews.
- **Medical diagnosis:** In medical applications, Naive Bayes can help in diagnosing diseases based on a set of symptoms or test results.
- **Credit Scoring:** In the finance industry, Naive Bayes can be used to assess the creditworthiness of individuals based on their financial history and other factors

REFERENCES

[1] Wikimedia Foundation. (2023, October 16). *Bayes' theorem*. Wikipedia. https://en.wikipedia.org/wiki/Bayes%27_theorem

[2] Comparison between multinomial and Bernoulli naïve Bayes for text ..., https://ieeexplore.ieee.org/document/8776800 (accessed Oct. 27, 2023).

[3] Banoula, M. (2023, August 29). *Naive Bayes classifier: Simplilearn*. Simplilearn.com. https://www.simplilearn.com/tutorials/machine-learning-tutorial/naive-bayes-classifier

[4]Coding ninjas studio. Available at: https://tinyurl.com/2s4x43pf

[5] https://www.upgrad.com/blog/gaussian-naive-bayes/

[6] Gaussian Naïve Bayes: https://www.ibm.com/topics/naive-bayes

REFERENCES

IMAGES

[1] https://tinyurl.com/5fsvthvr

[2] https://tinyurl.com/yckfb38n

[3] https://www.enjoyalgorithms.com/blog/naive-bayes-in-ml

[4] Naive Bayes classifier PPT. Available at: https://www.slideshare.net/knoldus/naive-bayes-classifier-99907521 (Accessed: 18 October 2023).

[5] Coding ninjas studio. Available at: https://www.codingninjas.com/studio/library/bernoulli-naive-bayes (Accessed: 25 October 2023).

[6] Naive Bayes PPT. Available at: https://www.slideshare.net/CloudxLab/naive-bayes-99709376 (Accessed: 18 October 2023).