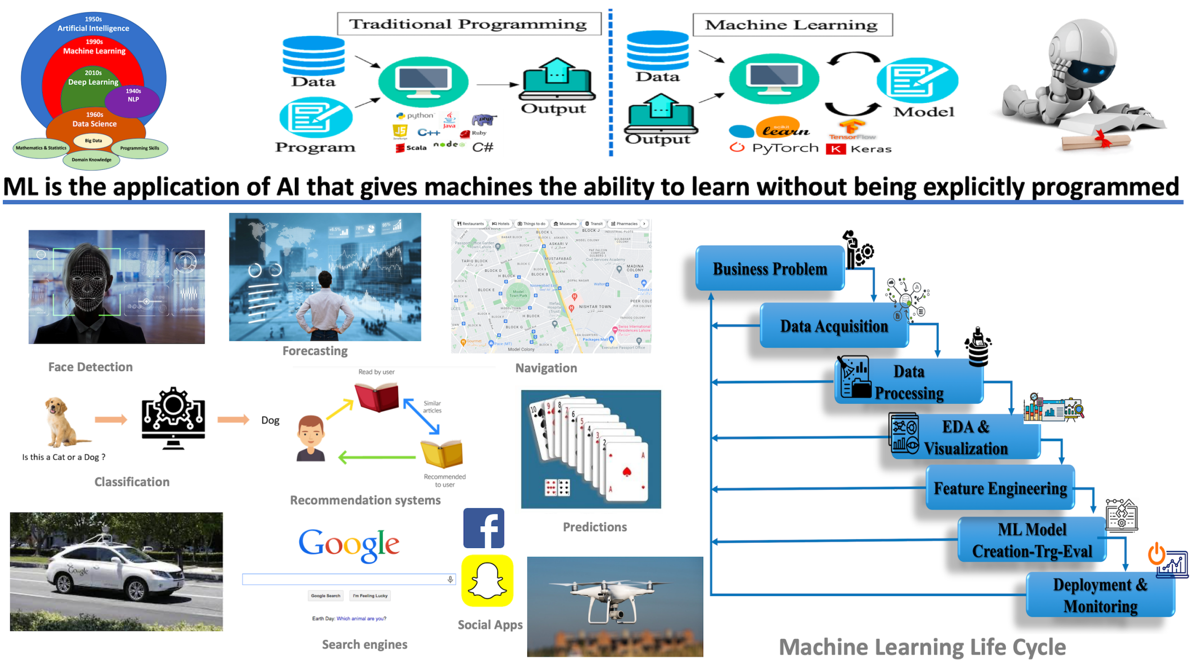
Department of Data Science

Course: Tools and Techniques for Data Science

Instructor: Muhammad Arif Butt, Ph.D.

Lecture 6.26 (Bayes Theorem and Naïve Bayes' Classifiers)

Open in Colab
[(https://colab.research.google.com/github/arifpucit/data-science/blob/master/Section-4-Mathematics-for-Data-Science/Lec-4.1(Descriptive-Statistics).ipynb)](https://colab.research.google.com/github/arifpucit/data-science/blob/master/Section-4-Mathematics-for-Data-Science/Lec-4.1(Descriptive-Statistics).ipynb)



Learning agenda of this notebook

- Bayes' Theorem
- Proof of Bayes' Theorem
- Formulas of Bayes' Theorem
- Examples of Bayes' Theorem
- Overview of Naïve Bayes' Classifier
- Naïve Bayes' Classifier for Datasets with Discrete Input Features
    - Example 1: Single Input Feature

- Example 2: Multiple Input Features
    - Example 3: Multiple Input Features
    - Example 4: Using Naïve Bayes' on Text Data
- Naïve Bayes' Classifier for Datasets with Continuous Input Features
    - Example 1: Single Input Feature
    - Example 2: Multiple Input Features
- Tasks to Do

1. Bayes' Theorem

- **Bayes' Theorem** (by Thomas Bayes), is a way of calculating a conditional probability without the joint probability.
- The theorem was introduced by Thomas Bayes, a statistician and philosopher and was first published in 1763.
- Bayes Theorem is the corner stone of Bayesian Learning methods that is a paradigm for constructing statistical models based on Bayes' Theorem

$$P(A \mid B) \quad = \quad \frac{P(A) \times P(B \mid A)}{P(B)}, \quad P(B) \neq 0$$

a. Proof of Bayes' Theorem

- To calculate the probability of event A to occur, given that event B has already occurred, we can use the following **Conditional Probability** formula in equation (i):

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad ------(i)$$

- Similarly, the probability of event B, given that event A has already occurred is given by equation2:

$$P(B \mid A) = \frac{P(B \cap A)}{P(A)} \qquad OR \qquad P(B \mid A) = \frac{P(A \cap B)}{P(A)} \qquad ------(ii)$$

- Multiplying both sides of equation (i) by $P(B)$ gives us:

$$P(A \mid B) * P(B) = P(A \cap B) \qquad ------(iii)$$

- Similarly, multiplying both sides of equation (ii) by $P(A)$ gives us:

$$P(B \mid A) * P(A) = P(A \cap B) \qquad ------(iv)$$

- Equating equations (iii) and (iv), we get:

$$P(A \mid B) * P(B) \quad = \quad P(B \mid A) * P(A)$$

$$P(A \mid B) \quad = \quad \frac{P(B \mid A) \times P(A)}{P(B)}, \quad P(B) \neq 0$$

b. Formulas of Bayes' Theorem

$$P(A \mid B) \quad = \quad \frac{P(A) \times P(B \mid A)}{P(B)}, \quad P(B) \neq 0$$

$$P(h \mid D) \quad = \quad \frac{P(h) \times P(D \mid h)}{P(D)}, \quad P(D) \neq 0$$

$$P(y = k \mid X) \quad = \quad \frac{P(y = k) \times P(X \mid y = k)}{P(X)},$$

- Where:

> - $P(y = k \mid X)$ is `Posterior Probability` : Probability of output label y given the input features X (also known as conditional probability).
> - $P(X|y = k)$ is `Liklihood` : Reverse of the posterior probability.
> - $P(y = k)$ is `Prior Probability` : Probability of an event that is calculated before considering the new information obtained.
> - $P(X)$ is also known as `evidence` or normalization constant.

c. Examples of Bayes' Theorem

- **Example 1:** What is the probability that a card drawn from a deck of playing cards is a Queen (Q), given that it is a card of Spades (S)?

$$P(y = Q \mid X) \quad = \quad \frac{P(y = Q) \times P(X \mid y = Q)}{P(X)}$$

$$P(y = Q) = \frac{4}{52}, \qquad P(X) = \frac{13}{52}, \qquad P(X \mid y = Q) = \frac{1}{4}$$

$$P(y = Q \mid X) \quad = \quad \frac{P(y = Q) * P(X \mid y = Q)}{P(X)} \quad = \quad \frac{\frac{4}{52} * \frac{1}{4}}{\frac{13}{52}} \quad = \quad \frac{1}{13}$$

- **Example 2:** In a school 60% of boys play football (F) and 36% of boys play cricket (C). The percentage of boys who play cricket given that they also play football is 40%. What is the percentage of those who play football given that they also play cricket?

$$P(F) = 0.6, \qquad P(C) = 0.36, \qquad P(C \mid F) = 0.4, \qquad P(F \mid C) = \ ?$$

$$P(F \mid C) = \frac{P(F) * P(C \mid F)}{P(C)}$$

$$P(F \mid C) = \frac{0.4 * 0.6}{0.36} = \frac{4}{6}$$

Example 3: Consider a factory where three machines (M1, M2, M3) are manufacturing 20, 30 and 50 items respectively per day. Moreover, the machines M1, M2, and M3 are manufacturing 5%, 3% and 1% defective items out of the total items they produce. Find the probability that an item selected at random from a batch is manufactured by M3 given that it is defective?

$$P(y = M1) = \frac{20}{100} = 0.2, \qquad P(y = M2) = \frac{30}{100} = 0.3, \qquad P(y = M$$

$$P(X \mid y = M1) = \frac{5}{100} = 0.05, \qquad P(X \mid y = M2) = \frac{3}{100} = 0.03, \qquad P(X \mid y =$$

$$P(y = M3 \mid X) = \frac{P(y = M3) \times P(X \mid y = M3)}{P(X)}$$

$$P(y = M3 \mid X) = \frac{P(y = M3) \times P(X \mid y = M3)}{P(X \cap M1) + P(X \cap M2) + P(X \cap M3)} \qquad \ldots.($$

We know that,

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)} \qquad \Longrightarrow \qquad P(A \cap B) = P(A \mid B) * P(B)$$

By substituting in equation (i)

$$P(y = M3 \mid X) = \frac{P(y = M3) \times P(X \mid y = M3)}{P(X \mid y = M1) * P(y = M1) + P(X \mid y = M2) * P(y = M2) -}$$

$$= \frac{0.5 * 0.01}{0.05 * 0.2 + 0.03 * 0.3 + 0.01 * 0.5} = 0.2083$$

2. Overview of Naïve Bayes' Classifier

The Naïve Bayes' (NB) is a probabilistic machine learning algorithm that use Bayes' Theorem for supervised machine learning classification.

- The Naïve Bayes' models are probabilistic classifiers, i.e., they not only assign a class label to a given sample, but they also provide an estimate of the probability that it belongs to that class. For example, a Naïve Bayes' model can predict that a given email has 80% chance of being a spam and 20% chance of being a ham.
  - `Deterministic classifiers` output a hard label for each sample, without providing probability estimates for the classes. Examples for such classifiers include K-nearest neighbors, decision trees, and SVMs.
  - `Probabilistic classifiers` output probability estimates for the k classes, and then assign a label to the sample based on these probabilities (typically the label of the class with the highest probability). Examples for such classifiers include Naïve Bayes' classifiers, logistic regression, and neural networks that use logistic/softmax in the output layer.
- The Naïve Bayes' classifier is called "naïve" because it assumes that the input variables/features are independent of each other, which is rarely the case. In other words, changing the value of one feature, does not directly change the value of any of the other features. This assumption is naïve because it is almost never true.
- For example, if you like to predict a cretain disease based on the body weight and blood pressure, the method assumes that these two variables are independent, which means that a person's blood pressure does not depend on his/her weight.
- Similarly, if you have temperature and humidity as input features, Naïve Bayes' assumes that temperature and humidity are independent of each other. That means changing the value of temperature does not directly change the value of humidity. However, in reality, we know that temperature and humidity are very much related to each other.
- Even though Naïve Bayes' is Naïve, it performs very well in applications, even when the features are not idependent of each other. Further more, compared to other ML algorithms, NB is fast, so it could be used for making predictions in real time

3. Naïve Bayes' Classifier for Datasets with Discrete Input Features

a. Example 1 (Single Input Feature)

- Given the training dataset, build a Naïve Bayes' Classifier, so that given a new instance of fruit having `green` color, you should be able to predict the category of the fruit being apple or orange.
- The idea is to compute the two probabilities, that is the probability of the fruit being apple or orange given its color. Which ever fruit type has the highest probability wins.
- Mathematically, we need to compute and later compare the following two conditional probabilities, and whichever, probability is greater, we predict the fruit accordingly.

$$P(y = apple \mid X = green) \quad \text{and} \quad P(y = orange \mid X = green)$$

- The Bayes' Theorem formula for predicting these probabilities is:

Ser	Color (X)	Fruit (y)
1	red	apple
2	red	apple
3	green	apple
4	green	orange
5	orange	orange
6	red	apple
7	green	apple
	...	...
	...	...

Step 1: Create `contingency or frequency table` for each feature of training data (Using Panda's `crosstab()` method). Since we have only one feature, which is color, so. we only need to build on frequency table as shown below:

X \ y	apple	orange	Total
red	33	0	33
green	20	7	27
orange	1	39	40
Total	54	46	100

Step 2: Compute the `marginal` probabilities for each of the output label (apple, orange):

$$P(y = apple) = \frac{54}{100} \qquad P(y = orange) = \frac{46}{100}$$

Step 3: Compute the `marginal` probabilities for the only input feature:

$$P(X = red) = \frac{33}{100} \qquad P(X = green) = \frac{27}{100}$$

$$P(X = orange) = \frac{40}{100}$$

Step 4: Compute the `conditional` probabilities:

$$P(X = green \mid y = apple) = \frac{20}{54}$$

$$P(X = green \mid y = orange) = \frac{7}{46}$$

Step 5: Substitute all the probabilities into the `Naïve Bayes' Formula`:

$$P(y = apple \mid X = green) = \frac{P(y=apple) \times P(X=green|y=apple)}{P(X=green)} =$$

$$P(y = orange \mid X = green) \quad = \quad \frac{P(y{=}orange) \times P(X{=}green \mid y{=}orange)}{P(X{=}green)} \quad =$$

Since

$$P(y = apple \mid X = green) \; > \; P(y = orange \mid X =$$

therefore, we classify the fruit given that its color is green is an apple

- In our case, the output variable (y) has only two outcomes (apple or orange). So we will compute the probability of all these features being apple as well as the probability of all these features being orange. Whatever, probability is larger is our output label.

- In step 5, the value of denominators of both formulas remain constant for a given input, i.e., $\frac{27}{100}$

- Since we compare these two probabilities, therefore, we can remove or ignore the term in the denominator. So for Naive Bayes' Classifier you can simplify the formula to

$$P(y \mid X) \quad \propto \quad P(y) \times P(X \mid y)$$

Zero Frequency Problem

- Suppose we are given a new instance of fruit having `red` color this time.
- In the frequency table, if you observe that out of 46 oranges you have zero orange having red color.
- Following the above described steps, we can use the Naïve Bayes' Classifier formula:

		y: apple	y: orange	Total
X	red	33	0	33
X	green	20	7	27
X	orange	1	39	40
	Total	54	46	100

$$P(y = apple \mid X = red) \quad = \quad P(y = apple) \times P(X = red \mid y = apple) \quad = \quad \cdot$$

$$P(y = orange \mid X = red) \quad = \quad P(y = orange) \times P(X = red \mid y = orange) \quad = \quad \frac{46}{100}$$

Since

$$P(y = apple \mid X = green) \; > \; P(y = orange \mid X =$$

therefore, we classify the fruit given that its color is red is an apple

> This seems to be working fine, but if we have many input features, the entire probability will become zero because one of the feature value is zero. This is called `Zero Frequency` problem, and it needs to be avoided. We can use Laplace smoothing to solve the problem of zero probability.

Laplace Smoothing

- **Laplace Smoothing** is usually applied by adding one count to the numerator and adding number of possible values of features to the denominator. In this example number of possible values of the only input feature is 3 (red, green, orange). So the probability after Laplace smoothing is:

$$P(y = apple \mid X = red) \quad = \quad P(y = apple) \times P(X = red \mid y = apple) \quad =$$

$$P(y = orange \mid X = red) \quad = \quad P(y = orange) \times P(X = red \mid y = orange) \quad = \quad \frac{46}{100}$$

Even after Laplace Smoothing, we still classify the fruit having red color as apple

b. Example 2 (Multiple Input Features)

	X			y
No	**Outlook (x1)**	**Temperature (x2)**	**Humidity (x3)**	**Rain**
1	Cloudy	Cool	High	yes
2	Sunny	Mild	High	no
3	Cloudy	Hot	Normal	no
4	Sunny	Cool	Normal	no
5	Sunny	Hot	Low	no
6	Cloudy	Mild	Normal	yes
7	Cloudy	Hot	High	yes
8	Cloudy	Cool	Low	no
9	Sunny	Cool	High	yes
10	Sunny	Hot	High	no
11	Cloudy	Hot	Low	no
12	Cloudy	Cool	Normal	yes
13	Sunny	Mild	Low	no
14	Cloudy	Cool	Low	yes
15	Sunny	Mild	Low	no

$$P(y = k \mid X_i) \quad \propto \quad P(y = k) \times P(X_i \mid y = k)$$

$$P(y = k \mid x_1, x_2, x_3, \ldots x_m) \quad \propto \quad P(y = k) * P(x_1 \mid y = k) *$$

$$P(y = k \mid x_1, x_2, x_3, \ldots x_m) \quad \propto \quad P(y = k) * \prod_{i=1}^{m} P(x_i \mid y = k)$$

$$P(y = k \mid x_1, x_2, x_3, \ldots x_m) \quad = \quad \hat{y} \quad = \quad argmax_k P(y = k) *$$

- **Let us assume that on a particular day the outlook is `coludy` , temperature is `cool` and humidity is `normal` .**
- Given the training dataset, build a Naïve Bayes' Classifier, so that given a new set of input features we should be able to predict if it will rain on that day?
- Mathematically, we need to compute and later compare the following two conditional probabilities, and whichever, probability is greater, we predict accordingly.

	X			y
No	**Outlook (x1)**	**Temperature (x2)**	**Humidity (x3)**	**Rain**
1	Cloudy	Cool	High	yes
2	Sunny	Mild	High	no
3	Cloudy	Hot	Normal	no
4	Sunny	Cool	Normal	no
5	Sunny	Hot	Low	no
6	Cloudy	Mild	Normal	yes
7	Cloudy	Hot	High	yes
8	Cloudy	Cool	Low	no
9	Sunny	Cool	High	yes
10	Sunny	Hot	High	no
11	Cloudy	Hot	Low	no
12	Cloudy	Cool	Normal	yes
13	Sunny	Mild	Low	no
14	Cloudy	Cool	Low	yes
15	Sunny	Mild	Low	no

$$P(y = yes \mid x_1 = cloudy, x_2 = cool, x_3 = normal)$$

$$P(y = no \mid x_1 = cloudy, x_2 = cool, x_3 = normal)$$

Step 1: Create contingency or frequency table for each feature of training data (Using Panda's `crosstab()` method). Since we have three features (outlook, temperature and humidity), so we need to construct three frequency tables as shown below:

Contingency Table for Outlook

x_1 \ y	yes	no	Total
Cloudy	5	3	8
Sunny	1	6	7
Total	6	9	15

Contingency Table for Temperature

x2 \ y	yes	no	Total
Cool	4	2	6
Mild	1	3	4
Hot	1	4	5
Total	6	9	15

Contingency Table for Humidity

x3 \ y	yes	no	Total
Low	1	5	6
Normal	2	2	4
High	3	2	5
Total	6	9	15

Step 2: Compute the probabilities for each of the output label (yes, no):

$$P(y = yes) = \frac{6}{15} \qquad\qquad P(y = no) = \frac{9}{15}$$

Step 3: Compute the conditional probabilities:

$$P(x_1 = cloudy \mid y = yes) \quad = \quad \frac{5}{6}$$

$$P(x_2 = cool \mid y = yes) \quad = \quad \frac{4}{6}$$

$$P(x_3 = normal \mid y = yes) \quad = \quad \frac{2}{6}$$

$$P(x_1 = cloudy \mid y = no) \quad = \quad \frac{3}{9} \qquad\qquad P(x_2 = cool \mid y = no) \quad = \quad \frac{2}{9}$$

$$P(x_3 = normal \mid y = no) \quad = \quad \frac{2}{9}$$

Step 4: Substitute all the probabilities into the Naïve Bayes' Formula:

$$P(y = yes \mid x_1 = cloudy, x_2 = cool, x_3 = normal)$$

$$= \quad P(y = yes) * P(x_1 = cloudy|y = yes) * P(x_2 = cool|y = yes) * P($$

$$= \quad \frac{6}{15} * \frac{5}{6} * \frac{4}{6} * \frac{2}{6} \quad = \quad \frac{240}{3240} \quad = \quad 0.074$$

$$P(y = no \mid x_1 = cloudy, x_2 = cool, x_3 = normal)$$

$$= \quad P(y = no) * P(x_1 = cloudy|y = no) * P(x_2 = cool|y = no) * P(x_3$$

$$= \quad \frac{9}{15} * \frac{3}{9} * \frac{2}{9} * \frac{2}{9} \quad = \quad \frac{108}{10935} \quad = \quad 0.009$$

$Since,$

$P(y = yes \mid x_1 = cloudy, x_2 = cool, x_3 = normal) \quad >$
$P(y = no \mid x_1 = cloudy, x_2 = cool, x_3 = normal)$

therefore, we predict that it will rain

c. Example 3 (Multiple Input Features)

In [1]:
```
import pandas as pd
df = pd.read_csv('datasets/titanic2.csv')
df
```

Out[1]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	who	adult_ma
0	0	3	male	22.0	1	0	7.2500	S	Third	man	Tr
1	1	1	female	38.0	1	0	71.2833	C	First	woman	Fal
2	1	3	female	26.0	0	0	7.9250	S	Third	woman	Fal
3	1	1	female	35.0	1	0	53.1000	S	First	woman	Fal
4	0	3	male	35.0	0	0	8.0500	S	Third	man	Tr
...	...	...	...	...	...	...	...	...	...	...	
886	0	2	male	27.0	0	0	13.0000	S	Second	man	Tr
887	1	1	female	19.0	0	0	30.0000	S	First	woman	Fal
888	0	3	female	NaN	1	2	23.4500	S	Third	woman	Fal
889	1	1	male	26.0	0	0	30.0000	C	First	man	Tr
890	0	3	male	32.0	0	0	7.7500	Q	Third	man	Tr

891 rows × 15 columns

In [2]:
```python
df1 = pd.concat([df['pclass'], df['sex'], df['survived']], axis=1)
df1
```

Out[2]:

	pclass	sex	survived
0	3	male	0
1	1	female	1
2	3	female	1
3	1	female	1
4	3	male	0
...	...	...	...
886	2	male	0
887	1	female	1
888	3	female	0
889	1	male	1
890	3	male	0

891 rows × 3 columns

Suppose we select a passenger at random from Titanic dataset, given that the passenger is female and belong to pclass 1, determine whether the passenger survived or not?

Create Contingency Tables

In [3]:
```python
pd.crosstab(index=df1['sex'], columns=df1['survived'], margins=True)
```

Out[3]:

survived	0	1	All
sex			
female	81	233	314
male	468	109	577
All	549	342	891

In [4]:
```python
pd.crosstab(index=df1['pclass'], columns=df1['survived'], margins=True)
```

Out[4]:

survived	0	1	All
pclass			
1	80	136	216
2	97	87	184
3	372	119	491
All	549	342	891

Contingency Table for sex

	y		
	not survived (0)	survived (1)	Total
female	81	233	314
male	468	109	577
Total	549	342	891

Contingency Table for pclass

	y		
	not survived (0)	survived (1)	Total
pclass = 1	80	136	216
pclass = 2	97	87	184
pclass = 3	371	119	491
Total	549	342	891

$$P(y = survived \mid x_1 = female, x_2 = pclass1)$$

$$= \quad P(y = survived) \quad * \quad P(x_1 = female | y = survived) \quad * \quad P(x_2 =$$

$$= \quad \frac{342}{891} * \frac{233}{342} * \frac{136}{342} \quad = \quad 0.1039$$

$$P(y = dead \mid x_1 = female, x_2 = pclass1)$$

$$= \quad P(y = dead) \quad * \quad P(x_1 = female | y = dead) \quad * \quad P(x_2 = pclass1|$$

$$= \quad \frac{549}{891} * \frac{81}{549} * \frac{80}{549} \quad = \quad 0.0132$$

$Since,$

$$P(y = survived \mid x_1 = female, x_2 = pclass1) \quad > \quad P(y = dead \mid x_1 = female, x_2 = pclass1)$$

therefore, we predict that passenger will survive

d. Example 4 (Text Data)

Document	Label
I am happy because I am playing cricket.	Positive
I am happy, not sad.	Positive
I am sad, I am not playing cricket.	Negative
I am sad, not happy.	Negative

- **Given a new sentence `I am playing cricket and happy`, Can you classify it as `Positive` or `Negative`.**

Step 1: Construct the contingency table containing frequency count for all the words:

		y	
	Positive (+)	Negative (-)	Total
I	3	3	6
am	3	3	6
happy	2	1	3
because	1	0	1
playing	1	1	2
cricket	1	1	2
sad	1	2	3
not	1	2	3
Total	13	13	26

$$P(y = + \mid x_1 = I, x_2 = am, x_3 = playing, x_4 = cricket, x_5 = happy)$$

Step 2: Compute the probabilities for each of the output label (Positive, Negative):

$$P(y = +) = \frac{1}{2} \qquad P(y = -) = \frac{1}{2}$$

Step 3: Compute the conditional probabilities:

$$P(x_1 = I \mid y = +) = \frac{3}{13} \qquad P(x_2 = am \mid y = +) = \frac{3}{13}$$
$$P(x_3 = playing \mid y = +) = \frac{1}{13}$$
$$P(x_4 = cricket \mid y = +) = \frac{1}{13} \qquad P(x_5 = happy \mid y = +) = \frac{2}{13}$$

$$P(x_1 = I \mid y = -) = \frac{3}{13} \qquad P(x_2 = am \mid y = -) = \frac{3}{13}$$
$$P(x_3 = playing \mid y = -) = \frac{1}{13}$$
$$P(x_4 = cricket \mid y = -) = \frac{1}{13} \qquad P(x_5 = happy \mid y = -) = \frac{1}{13}$$

Step 4: Substitute all the probabilities into the Naïve Bayes' Formula:

$$P(y = + \mid x_1 = I, x_2 = am, x_3 = playing, x_4 = cricket, x_5 = h$$

$$= P(y = +) * P(x_1 = I|+) * P(x_2 = am|y = +) * P(x_2 = playing|y = +) *$$

$$= \frac{1}{2} * \frac{3}{13} * \frac{3}{13} * \frac{1}{13} * \frac{1}{13} * \frac{2}{13} = 0.000024$$

$$P(y = - \mid x_1 = I, x_2 = am, x_3 = playing, x_4 = cricket, x_5 = h$$

$$= \quad P(y = -) \ * \ P(x_1 = I|-) \ * \ P(x_2 = am|y = -) \ * \ P(x_2 = playing|y = -) \ * $$

$$= \quad \frac{1}{2} * \frac{3}{13} * \frac{3}{13} * \frac{1}{13} * \frac{1}{13} * \frac{1}{13} \quad = \quad 0.0000121$$

$$Since,$$

$$P(y = + \mid x_1 = I, x_2 = am, \cdots x_5 = happy) \quad >$$

$$P(y = - \mid x_1 = I, x_2 = am, \cdots x_5 = happy)$$

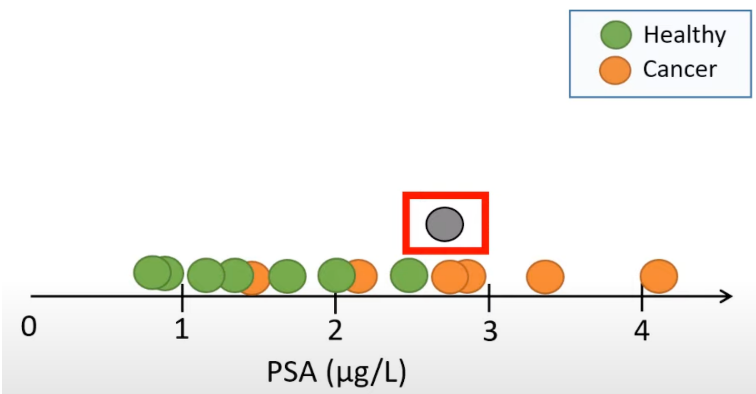
4. Naïve Bayes' Classifier for Datasets with Continuous Input Features

a. Example 1 (Single Input Feature)

- Consider this hypothetical data set showing the measures of the Prostate-Specific Antigen (PSA) concerntration from blood samples of fourteen persons, out of which seven are patients with prostate cancer while seven are healthy.
- Dataset has one input feature that is PSA level which is a continuous variable, while the output label has two possible values `Cancer` and `Healthy`, and the objective is to build a Naïve Bayes Classifier to predict if a given patient suffers from cancer or not given his/her PSA value.
- `100$ Question` : How to compute the probabilities when X is a continuous variable?
- `Answer` : Use **Gaussian Naïve Bayes** model, which assumes that the continuous input feature variables follow a Gaussian (Normal) distribution.

No	PSA	Status
1	4.1	Cancer
2	3.4	Cancer
3	2.9	Cancer
4	2.8	Cancer
5	2.7	Cancer
6	2.1	Cancer
7	1.6	Cancer
8	2.5	Healthy
9	2.0	Healthy
10	1.7	Healthy
11	1.4	Healthy
12	1.2	Healthy
13	0.9	Healthy
14	0.8	Healthy

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \, e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



- Suppose we like to predict if a person with `PSA level of 2.6` ug/L

Calculate Sample Mean and Standard Deviation of Healthy group:

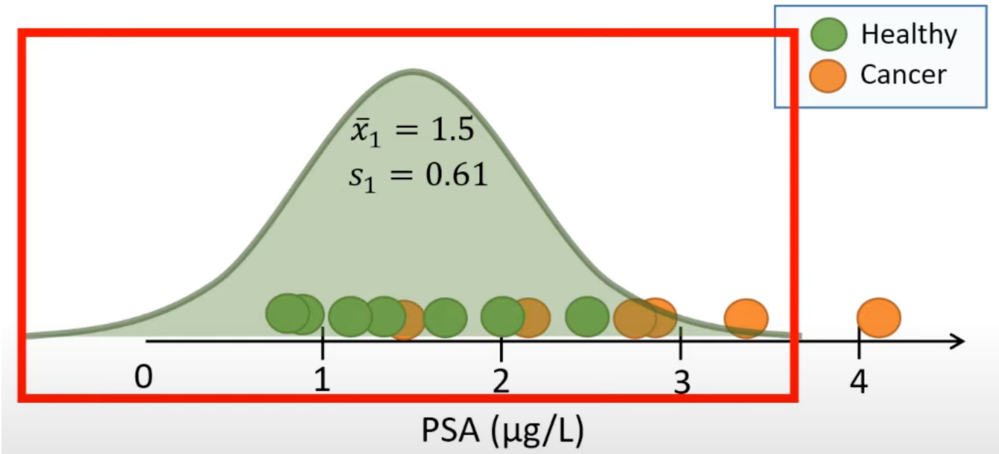
- `Healthy Group:`

$$\mu_{healthy} = \frac{\sum_{i=1}^{n} x_i}{n} = 1.5$$

$$\sigma_{healthy} = \sqrt{\frac{\sum (x_i - \mu)^2}{n-1}} = 0.61$$

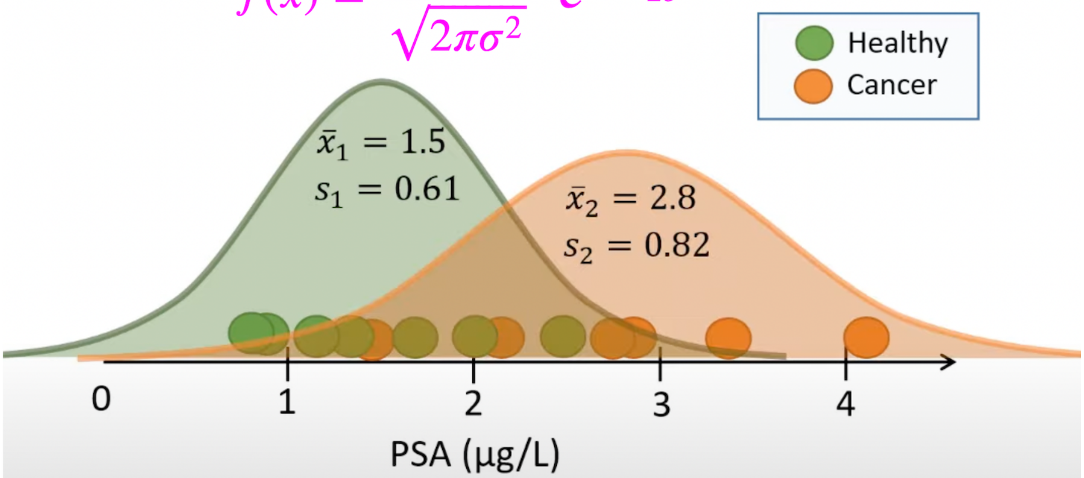
- If we plug in the $\mu_{healthy}$ and $\sigma_{healthy}$ values for the healthy group with different x_i values of PSA in the probability density function of Gaussian Distribution, we get different values which if plotted will give the following bell shaped curve for Healthy persons:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



Calculate Sample Mean and Standard Deviation of all Groups:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$



- `Healthy Group:` $\mu_{healthy} = 1.5,$ $\sigma_{healthy} = 0.61$

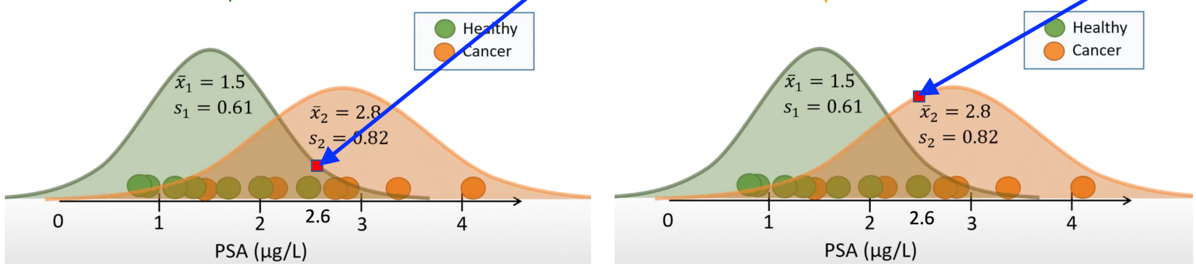
- `Cancer Group:` $\mu_{cancer} = 2.8,$ $\sigma_{cancer} = 0.82$

- Now we need to calculate $f(psa = 2.6 \mid Healthy)$, and $f(psa = 2.6 \mid Cancer)$ for which we will use the probability density function of Gaussian Distribution as shown below:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

$$f(psa = 2.6 \mid Healthy) = \frac{1}{\sqrt{2 \times 3.146 \times 0.61^2}} e^{-\frac{(2.6-1.5)^2}{2 \times 0.61^2}} = 0.13$$

$$f(psa = 2.6 \mid Cancer) = \frac{1}{\sqrt{2 \times 3.146 \times 0.82^2}} e^{-\frac{(2.6-2.8)^2}{2 \times 0.82^2}} = 0.47$$



> These values 0.13 and 0.47 are not the probabilities, rather are the values of the probability density function for the height of these curves where PSA=2.6

- Substitute these values in the Naïve Bayes' Formula:

$$P(y = Healthy \mid psa = 2.6) \quad = \quad P(y = Healthy) * P(psa = 2.6 | y = Healthy)$$

$$P(y = Cancer \mid psa = 2.6) \quad = \quad P(y = Cancer) * P(psa = 2.6 | y = Cancer) \quad =$$

$$Since,$$
$$P(y = Cancer \mid psa = 2.6) \quad > \quad P(y = Healthy \mid psa = 2.6)$$

therefore, we classify the person has Cancer.

b. Example 2 (Multiple Input Features)

- Consider this hypothetical dataset having two input features `PSA` level and `age`.

Calculate Sample Mean and Standard Deviation of all Groups:

- `Cancer Group:`
  $\mu_{psa} = 2.8,$ $\sigma_{psa} = 0.82$

No	PSA	Age	Status
1	4.1	78.0	Cancer
2	3.4	70.0	Cancer
3	2.9	62.0	Cancer
4	2.8	66.0	Cancer
5	2.7	70.0	Cancer
6	2.1	65.0	Cancer
7	1.6	58.0	Cancer
8	2.5	68.0	Healthy
9	2.0	64.0	Healthy
10	1.7	62.0	Healthy
11	1.4	70.0	Healthy
12	1.2	72.0	Healthy
13	0.9	67.0	Healthy
14	0.8	59.0	Healthy

$\mu_{age} = 67, \qquad \sigma_{age} = 6.45$

- `Healthy Group:`

$\mu_{psa} = 1.5, \qquad \sigma_{psa} = 0.61$

$\mu_{age} = 66, \qquad \sigma_{age} = 4.58$

Calculate the conditional probabilities:

- `Cancer Group:`

$$f(psa = 2.6 \mid Cancer) = \frac{1}{\sqrt{2\pi\sigma^2}} \; e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2 \times 3.146 \times 0.82^2}} \; e^{-\frac{(x-2.8)^2}{2 \times 0.82^2}}$$

$$f(age = 70 \mid Cancer) = \frac{1}{\sqrt{2\pi\sigma^2}} \; e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2 \times 3.146 \times 6.45^2}} \; e^{-\frac{(x-67)^2}{2 \times 6.45^2}}$$

- `Healthy Group:`

$$f(psa = 2.6 \mid Healthy) = \frac{1}{\sqrt{2\pi\sigma^2}} \; e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2 \times 3.146 \times 0.61^2}} \; e^{-\frac{(x-1.5)^2}{2 \times 0.61^2}}$$

$$f(age = 70 \mid Healthy) = \frac{1}{\sqrt{2\pi\sigma^2}} \; e^{-\frac{(x-\mu)^2}{2\sigma^2}} = \frac{1}{\sqrt{2 \times 3.146 \times 4.58^2}} \; e^{-\frac{(x-66)^2}{2 \times 4.58^2}}$$

- Substitute all the probabilities into the Naïve Bayes' Formula:

$P(y = Cancer \mid psa = 2.6, age = 70) \quad = \quad P(y = Cancer) * P(psa = 2.6 | y =$

$$= \quad \frac{1}{2} * 0.47 * 0.055 \quad = \quad 0.013$$

$$P(y = Healthy \mid psa = 2.6, age = 70) \quad = \quad P(y = Healthy) * P(psa = 2.6 | y =$$
$$= \quad \frac{1}{2} * 0.13 * 0.059 \quad = \quad 0.004$$

$$Since,$$
$$P(y = Cancer \mid psa = 2.6, age = 70) \quad >$$
$$P(y = Healthy \mid psa = 2.6, age = 70)$$

therefore, we classify the person has Cancer.

Tasks To Do

Task 1:

- **Given the dataset, use Naïve Bayes' Classifier and predict if the new instance having values A=0, B=1, C=0, belong to positive class or negative class.**

No	A	B	C	Class
1	0	0	0	Positive
2	0	0	1	Negative
3	0	1	1	Negative
4	0	1	1	Negative
5	0	0	1	Positive
6	1	0	1	Positive
7	1	0	1	Negative
8	1	0	1	Negative
9	1	1	1	Positive
10	1	0	1	Positive

Task 2:

- **Given the dataset, use Gaussian Naïve Bayes' Classifier and predict the gender of a person having height 6 feet, weight 130 lbs, and foot size 8 inches.**

No	Height (ft)	Weight (lbs)	Foot Size (inches)	Gender
1	6.00	180	12	Male
2	5.92	190	11	Male
3	5.58	170	12	Male
4	5.92	165	10	Male
5	5.00	100	6	Female
6	5.50	150	8	Female
7	5.42	130	7	Female

Task 3:

- **Given the dataset, use Naïve Bayes' Classifier and classify a new observation: Outlook = `Overcast` , Temp = `60` , Humidity= `62` , Wind= `False` , into either one of the two classes Yes or No.**

Day	Outlook	Temp	Humidity	Wind	Play Tennis
D1	Sunny	85	85	FALSE	No
D2	Sunny	80	90	TRUE	No
D3	Overcast	83	86	FALSE	Yes
D4	Rainy	70	96	FALSE	Yes
D5	Rainy	68	80	FALSE	Yes
D6	Rainy	65	70	TRUE	No
D7	Overcast	64	65	TRUE	Yes
D8	Sunny	72	95	FALSE	No
D9	Sunny	69	70	FALSE	Yes
D10	Rainy	75	80	FALSE	Yes
D11	Sunny	75	70	TRUE	Yes
D12	Overcast	72	90	TRUE	Yes
D13	Overcast	81	75	FALSE	Yes
D14	Rainy	71	91	TRUE	No