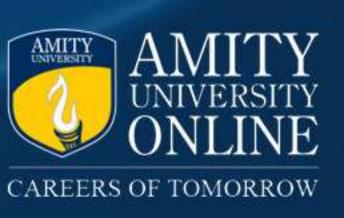


However, class variable should always be categorical in nature.



# NAIVE BAYES FOR CLASSIFICATION X

Naive Bayes - Estimating probabilities from Data - Example illustration 2

Consider data set below. Where indicator variables are mixture of data types and objective is to predict Gender of an

Class variable

unknown person.

**Numerical variable** 

Age	Weight	Gender
32	Heavy	Male
39	Normal	Female
21	Heavy	Male
22	Heavy	Female
31	Normal	Male

Indicator variable

categorical variable

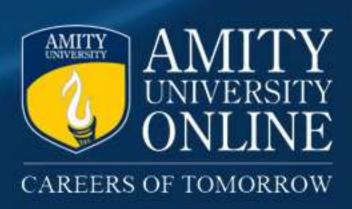
Figure 68: Hypothetical data set

Suppose we are to predict Gender class of an observation:

( Age = 33, Weight = Heavy), using Naive Bayes classifier using data set above.

For continuous indicator variables present in the data set, we use Normal distribution to compute the probability density function as defined below.

$$P(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{\frac{(x-\mu)}{2\sigma^2}} \qquad (33)$$
 where, 
$$\mu = mean \qquad \sigma = standard deviation$$



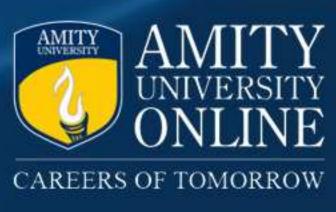
# NAIVE BAYES FOR CLASSIFICATION XI

Naive Bayes - Estimating probabilities from Data - Example illustration 2

Objective: To predict class of Gender for given input: (Age = 33, Weight = Heavy)

In order to predict the class, we need to compute following two probabilities using Naive Bayes:

- 1. P (Gender = Male | Age = 33, Weight = Heavy)
- 2. P (Gender = Female | Age = 33, Weight = Heavy)



## NAIVE BAYES FOR CLASSIFICATION XII

### Naive Bayes - Estimating probabilities from Data - Example illustration 2

$$P(Age = 33 \mid Gender = Male) \times P(Weight = Heavy \times Gender = Male) \times P(Gender = Male)$$

#### Using data set,

P (Gender = Male) = 
$$3/5 = 0.6$$

P (Weight = Heavy | Gender = Female) = 
$$2/3 = 0.66$$

			n		8
$Q^{x} =$	~ /	$\frac{1}{n-1}$	$\Sigma$	( x <sub>i</sub> –	$\mu_{x})^{2}$
		Server Thinks	i = 1		

 $\mu_{x} = \frac{1}{n} \sum_{i} x_{i}$ 

Age	Weight	Gender
32	Heavy	Male
39	Normal	Female
21	Heavy	Male
22	Heavy	Female
31	Normal	Male

Figure 68: Hypothetical data set

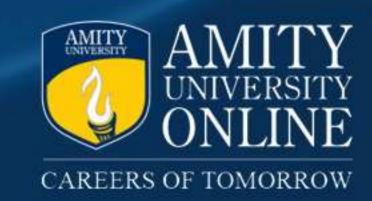
Age	Weight	Gender
32	Heavy	Male
21	Heavy	Male
31	Normal	Male

### In order to compute above:

- 1. Find Age of all data points with Gender = Male.
- 2. Compute Mean and Standard deviation of Age for data set in right. On calculating we have,  $\mu$  = 28 and  $\sigma$  = 6.08.

P (Age = 33 × Gender = Male) = 
$$\frac{1}{\sqrt{2 \times 3.14 \times (6.08)^2}} e^{\frac{(33-28)^2}{2 \times 6.08^2}} = 0.08$$

4. P (Gender = Male | Age = 33, Weight = Heavy) = 
$$\frac{0.08 \times 0.66 \times 0.6}{P \text{ (Age = 33 | Weight = Heavy)}} = 0.024$$



## NAIVE BAYES FOR CLASSIFICATION XIII

### Naive Bayes- Estimating probabilities from Data- Example illustration 2

P (Gender = Female | Age = 33, Weight = Heavy)

$$P(Age = 33 \mid Gender = Female) \times P(Weight = Heavy \times Gender = Female) \times P(Gender = Female)$$

C	nn	data	COT
		uala	36.

P (Gender = Female) = 2/5 = 0.4

P (Weight = Heavy | Gender = Female) = 1/2 = 0.50

P(Age = 33 | Gender = Female) = ?

Age	Weight	Gender
32	Heavy	Male
39	Normal	Female
21	Heavy	Male
22	Heavy	Female
31	Normal	Male

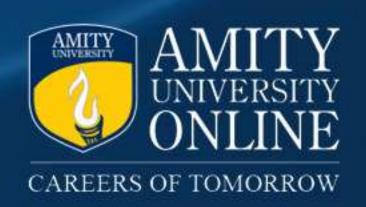
Age	Weight	Gender
39	Normal	Female
22	Heavy	Female

#### In order to compute above:

- Find Age of all data points with Gender = Female.
- 2. Compute Mean and Standard deviation of Age for data set in right. On calculating we have,  $\mu = 27$  and  $\sigma = 7.07$ .

P (Age = 33 × Gender = Female) = 
$$\frac{1}{\sqrt{2 \times 3.14 \times (7.07)^2}} e^{\frac{(33-27)^2}{2 \times 7.07^2}} = 0.063$$

4. P (Gender = Female | Age = 33, Weight = Heavy) = 
$$\frac{0.063 \times 0.5 \times 0.4}{P \text{ (Age = 33 \times Weight = Heavy)}} = 0.0126$$



# NAIVE BAYES FOR CLASSIFICATION XIV

### Naive Bayes- Estimating probabilities from Data- Example illustration 2

Age	Weight	Gender
32	Heavy	Male
39	Normal	Female
21	Heavy	Male
22	Heavy	Female
31	Normal	Male

P(Gender = Male | Age = 33, Weight = Heavy) > P(Gender = Female | Age = 33, Weight = Heavy)

Hence, a new data point (Age = 33, Weight = Heavy) is predicted as Male.

