NAIVE BAYES FOR CLASSIFICATION I

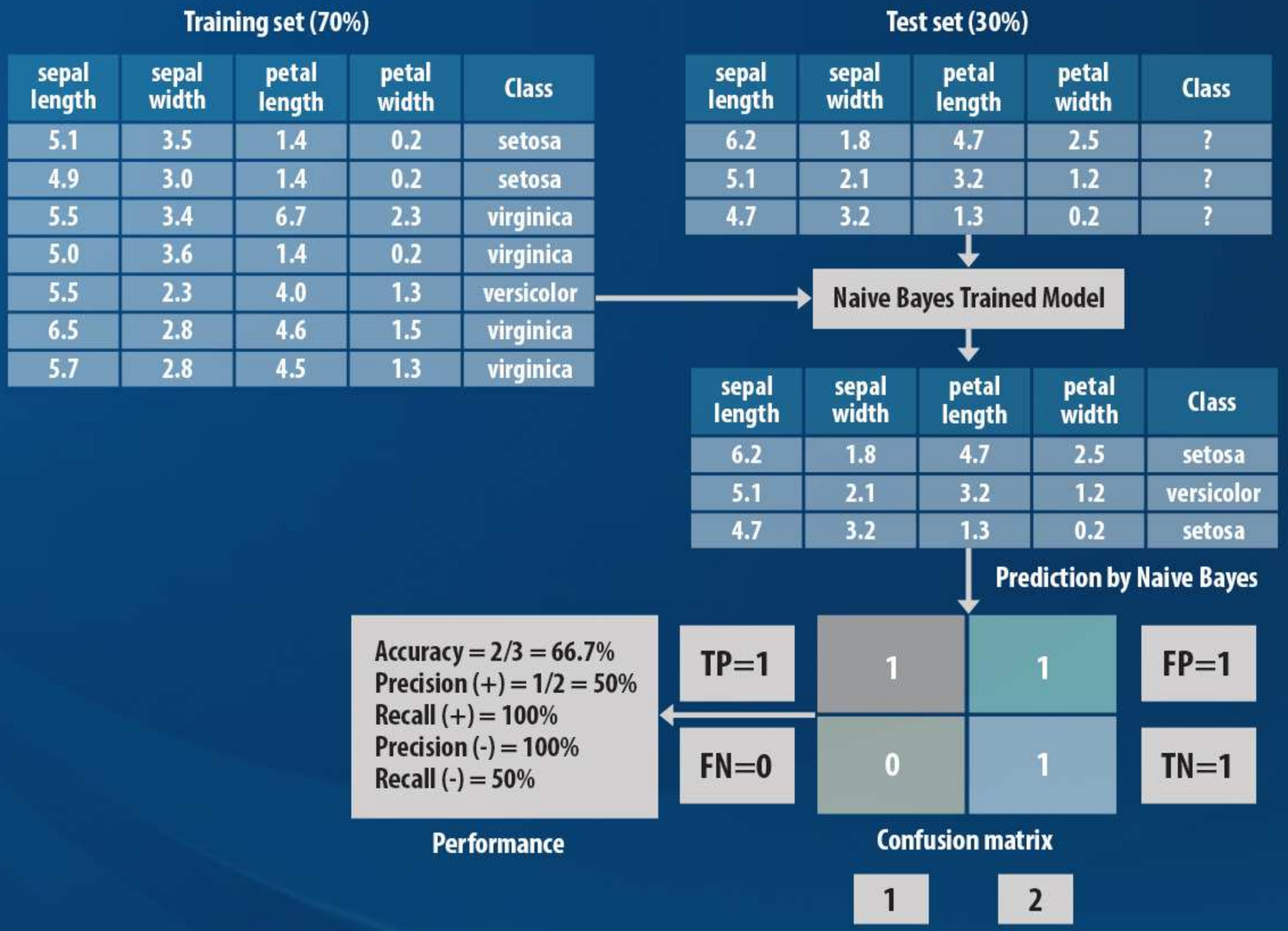

Figure 69: Naive Bayes Classifier

To avoid any probabilistic query resulting in zero, a Laplace Correction is introduced.

ISSUES WITH NAIVE BAYES CLASSIFIER I

For given data set with $X = (x_1, x_2, x_3, \ldots, x_n, C)$ features. Where C is the class label.
The Naive Bayes classifier uses Equation 35 to estimate the class of new test point.

$$P(C \mid x_1, x_2, x_3, \ldots, x_n) = \frac{P(x_1|C) \times P(x_2|C) \times P(x_3|C), ., ., x\, P(x_n|C) \times P(C)}{P(x_1, x_2, x_3, \ldots., x_n)} \quad (34)$$

The major issue is that If one of the conditional probabilities (right hand side of Equation) is zero, then the entire expression becomes zero. Thereby making Naive Bayes classifier failing to predict.

The solution to the problem is to estimate conditional probability using Laplace correction.

$$P(x_i \mid C) = \frac{N_{ic}+1}{N_c+c} \quad (35)$$

Where:

N_c : Number of instances in the class

c : number of classes

N_{ic} : Number of instances having feature value x_i in class C

ISSUES WITH NAIVE BAYES CLASSIFIER II

- **Laplace correction Example**

Consider data set in Figure 70. Let we are to predict the probability of test case:

$P(\text{Cancer} = \text{Yes} \mid \text{Smoke} = \text{No}, \text{X-Ray} = \text{Normal})$?

We find the probability of test case using Naive Bayes classifier as below,

Smoke	X-ray	Cancer
Yes	Normal	No
Yes	Abnormal	Yes
Yes	Normal	No
No	Abnormal	No
No	Abnormal	Yes

Figure 70: Hypothetical data set

$$P(\text{Cancer} = \text{Yes} \mid \text{Smoke} = \text{No}, \text{X-Ray} = \text{Normal}) = \frac{P(\text{Smoke} = \text{No} \mid \text{Cancer} = \text{Yes}) \times P(\text{X-Ray} = \text{Normal} \mid \text{Cancer} = \text{Yes}) \times P(\text{Cancer} = \text{Yes})}{P(\text{Smoke} = \text{No} \mid \text{X-Ray} = \text{Normal})} \quad (36)$$

Here the probability of expression $P(\text{X-Ray} = \text{Normal} \mid \text{Cancer} = \text{Yes})$ is 0 by which the prediction value becomes 0. We use laplace correction here:

$$P(\text{X-Ray} = \text{Normal} \mid \text{Cancer} = \text{Yes}) = \frac{0+1}{2+2} = 0.25\% \quad (37)$$

NAIVE BAYES (SUMMARY)

- Naive Bayes is the simplest, easy model to be used for any application.
- It is a white box model. Meaning that Naive Bayes not only predicts but Naive Bayes can also explain the reason to its prediction.
- Handle missing values by ignoring the instance during probability estimate calculations.
- Naive Bayes is not a good choice for domains where, features are correlated. It is for the reason that Naive Bayes assumes that all features are independent to each other given the class variable.