ESTIN

Machine Learning

S4 2023-2024

Lab (Naïve Bayes)

Exercise1:

1. Consider the dataset **breast_cancer** from **sklearn**. Split the data into training and testing sets (30% for testing set, random_state=0).
2. Train a **Naive Bayes classifier** (apply **GaussianNB**).
3. Compute the **score** of the model **(**for both training and testing).
4. Use the **predict_proba** method to find the probability, considering few test data sets.
5. Draw the **confusion matrix**.

Exercise2:

1. Load the **Iris dataset**. Split the data into training and testing sets (30% for testing set, random_state=0).
2. Pick the appropriate type of NBs given the nature of your data set. Choices are: MultinomialNB, BernoulliNB.
3. Why is BernoulliNB performing like it does?
4. Use **cross validation** to see how well your choice works.
5. Compare the cross validation scores for Bernoulli and multinomial Naive Bayes.
6. Suppose a sample has the following data:
   Sepal_length=5.85, sepale_with=4.5, petal_lengh=1, petal_with=2.5. Find its class using 1) MultinomialNB; 2) BernoulliNB.