Homework 1 Simulation Exercises

Exercise 1

The goal of this exercise is to learn and implement the Naive Bayes Classifier.

Part A: What is the main difference between Naive Bayes and Optimal Bayes classifiers? Explain what points we lose when using Naive Bayes, and when using it is reasonable.

Part B: The 'Penguins' dataset is attached in the HW folder. First, do any kind of preprocessing you think is required (normalizing, handling missing values, distribution visualizing, etc.) Explain each preprocessing you employ.

Part C: The dataset contains three classes. Implement a Naive Bayes Classifier from scratch (do not use Tensorflow and predefined modules). Use one-vs-all technique when classifying the test set. Display the confusion matrix and evaluation parameters (accuracy, precision, recall).

Exercise 2

A data set named 'image' is attached in the folder. It contains images of two classes: jungle and sea. Design a heuristic decision boundary based on the features of images to classify them. Mention the samples which have been misclassified and explain the reason of misclassifications regarding the decision boundary.