



Machine Learning a.y. 22-23

Homework 1: Report

Davide Gabrielli - gabrielli.1883616@studenti.uniroma1.it

December 8, 2022

1 Introduction

In this report we will discuss the design and implementation for the given two tasks about classification and regression about collisions of drones in a 2D environment.

1.1 Dataset

The dataset we will use is the one provided by the course. It contains 1000 samples of drone collisions in a 2D environment. The dataset contains the following features:

- **For 5 UAVs** (where i is the UAV number)):
 - **UAV_i_track**: clockwise angle from north between the i th UAV and its target ($0, 2\pi$)
 - **UAV_i_x**: x coordinate of the i th UAV in meters
 - **UAV_i_y**: y coordinate of the i th UAV in meters
 - **UAV_i_vx**: x velocity of the i th UAV in m/s
 - **UAV_i_vy**: y velocity of the i th UAV in m/s
 - **UAV_i_target_x**: x coordinate of the i th UAV target in meters
 - **UAV_i_target_y**: y coordinate of the i th UAV target in meters
- **num_collisions**: number of collisions in the sample
- **min_CPA**: minimum CPA in the sample (An estimated point in which the distance between two objects, of which at least one is in motion, will reach its minimum value)

The dataset is unbalanced, as it contains 1000 samples divided into 5 classes (0, 1, 2, 3, 4) that represent the number of collisions in the sample, as shown in (Fig. 1). This is also due to the fact that the distribution of the number of collisions is not uniform.

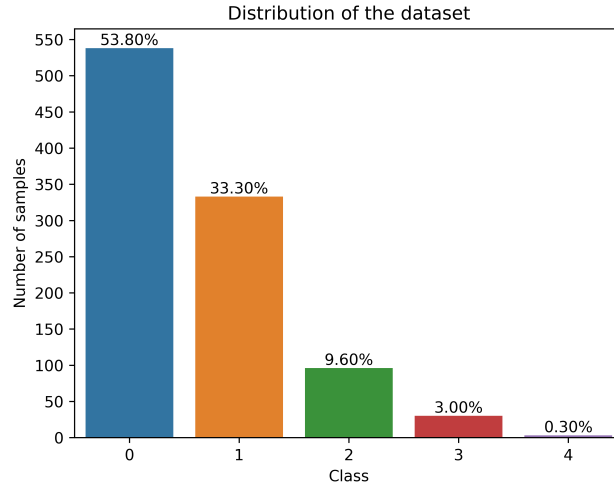


Figure 1: The dataset is composed of 1000 samples divided into 5 classes that represent the number of collisions.

1.2 Classification

The classification task is to predict the number of collisions in a sample. We will use several classification algorithms to compare their performance. The algorithms we will use are:

- **Logistic Regression**
- **Decision Tree**
- **Random Forest**
- **Support Vector Machine**
- **K-Nearest Neighbors**

Since the dataset is unbalanced, we expect better results from Random Forest and Support Vector Machine, as they are able to handle unbalanced datasets. We will use the following metrics to evaluate the performance of the algorithms:

- **Accuracy:** the number of correct predictions divided by the total number of predictions
- **Precision:** the number of true positives divided by the number of true positives plus the number of false positives
- **Recall:** the number of true positives divided by the number of true positives plus the number of false negatives
- **F1-Score:** the harmonic mean of precision and recall



Since the dataset is unbalanced, we will use the F1-Score as the main metric to evaluate the performance of the algorithms. We will also use the confusion matrix to visualize the performance of the algorithms.

1.3 Regression

The regression task is to predict the minimum CPA in a sample. We will use several regression algorithms to compare their performance. The algorithms we will use are:

- **Linear Regression**
- **Decision Tree**
- **Random Forest**
- **Support Vector Machine**
- **K-Nearest Neighbors**

We will use the following metrics to evaluate the performance of the algorithms:

- **Mean Absolute Error:** the average of the absolute differences between the predictions and the actual values
- **Mean Squared Error:** the average of the squared differences between the predictions and the actual values
- **Root Mean Squared Error:** the square root of the mean squared error

We will use the Root Mean Squared Error as the main metric to evaluate the performance of the algorithms.



2 Learning Model

Describe the state-of-the-art protocol you choose



3 Experiments

Discuss simulations settings and obtained results. We expect you to compare the obtained results with baseline approaches (geographic routing, random routing and your hw1 q-learning solution). Please, include comments and plots.



4 Conclusions

Summarize your results and your findings.