Assignment 5: Classification

July 16, 2022

[Submission URL](https://drive.google.com/drive/folders/1guTB3OZS7nXdDFxjP1AzvINilE8Wff-d?usp=sharing)

Submission Deadline: 18/07/2022, 2.30 PM

1. Pokemon dataset is a collection of information of pokemons from seven generations. Data can be loaded as dataframe using the command:

*data = pd.read\_csv('https://tinyurl.com/acm16072022')*.

Using this dataset answer the following questions.

* 1. Which features of the dataset are more correlated and carries redundant information? (Drop ‘name’ and ‘type1’ columns for answering the question). You can use scatter plot matrix (or) calculate the correlation coefficient among the features to answer the question.
  2. Split the dataset as train and test set. Test set should have 0.2 fraction of the total dataset.
  3. Use the train dataset to train five multi-class classifiers of your choice. Your predictors (or) X variables should not include ‘name’ and ‘type1’ columns in it. Take your response variable as ‘type1’.
  4. Report the best classifier based on the following metrics applied on test dataset:
     1. Accuracy
     2. F1-score
     3. AUC-ROC curve
     4. Precision-Recall curve
     5. True positivity rate (or) sensitivity