

Intelligent Systems Final Project - Toolbox

Group Members - Ketaki Tilak, Abhik Roy

Description

We have created a toolbox of 2 classification, 1 clustering, 1 regression algorithms, and a vectorization algorithm that can be used for ranking and retrieval . We have implemented Logistic Regression and Decision Tree (for classification) and Linear Regression (for regression), KMeans Clustering(clustering) and TF_IDF Matrix generation(for retrieval and ranking).

For logistic and Linear regression, we have implemented the fit() function which will fit the model to the input data; and we will also write a function to predict the target variable from a set of input features.

For the third model in our toolbox, we have implemented a decision tree classifier. For this as well, we have written the fit() and predict() functions with all the necessary helper methods.

This is a decision tree classifier for continuous numerical data.

We have implemented a KMeans Clustering model, with additional tools for cluster visualizations and to find the optimal value of K for a given dataset.

We have implemented tools to generate a TF-IDF Matrix from text data to perform tasks like ranking and retrieval using similarity metrics.

We have benchmarked the performance of all these models by comparing them to respective sklearn models.

We have used python for the module, all the files are in the directory statsbox and can be imported as **from statsbox import file**

Objectives

We have tried to complete following objectives:

1. Proposal (10 pts)
2. Demo (10 pts)
3. Defend Challenging (10 pts)
4. Tasks: Classification, Regression, Clustering, Retrieval, Ranking (5 pts)
5. Documentation (2 pts)
6. Benchmarking with baselines (5 pts)
7. High-performance implementation on CPU (5 pts)
8. Illegal input handling (2 pts)
9. Cross-Platform Compilation (Mac, Windows, Linux (Ubuntu)) (3 pts)
10. Non-Trivial Optimization(2) (2 points)

For a total of 54 points.