Advanced Methods in Data Science - DTSC104

Project

General and Milestones Description

Dr. Eng. Catherine M. Elias

Mechatronics Engineering department,
Faculty of Engineering and Materials Science (EMS), German University in Cairo (GUC)

Mini-Project

Requirements

Get your own data from real-life application (2-3 features)

- Cluster/label the data using GA and K-mean Approaches → Unsupervised Learning
- Compare between both algorithms results in terms of 2 KPIs.

Generate new unlabeled data

- Classify the new generated data using both SVM and ANN → Supervised Learning
- Compare between both algorithms results in terms of 2 KPIs.

Milestone 1

- In the first milestone, each team is requested to:
  1. Find the dataset
  2. Label/cluster the dataset using the GA and the K-Mean clustering algorithms based on at least 2 input features.
  3. Run the GA algorithm to find the final labeling/clustering
  4. Run the K-Mean clustering algorithm to find the final labeling/clustering
  5. Compare in terms of the MSE and the computational time KPIs

Mini-Project

Milestone 1

- You are requested to submit:
  1. The collected dataset,
  2. GA labeling Python code,
  3. K-mean clustering Python code, and
  4. presentation **5 slides only** with the progress in this milestone (include graphs).

- All the files should be zipped and uploaded to the form:

https://docs.google.com/forms/d/e/1FAIpQLSfA8qQ3bcgzAknUUDCdaWmjnexRFhvaw5Mx25puofLJ3FlY0Q/viewform?usp=sf_link

- The first milestone deadline is **Thursday 13th of April, 2023 at 11:59 PM**.

Milestone 2

- In the second milestone, each team is requested to:
  1. Implement the SVM code for new data prediction and Classification.
  2. Tune the parameters (manually) till reaching the best results.
  3. Observe the effect of the parameters changing on the performance in terms of accuracy, MSE execution time, prediction time.

Milestone 2

- You are requested to submit:
  1. SVM Classification Python code,
  2. presentation **5 slides only** with the progress in this milestone (include graphs).

- All the files should be zipped and uploaded to the form:

https://docs.google.com/forms/d/e/1FAIpQLSfYkMIaRFF-PK_n4F74XWNrJKiJtWryYFOd4zhP2bQwzHZoEA/viewform?usp=sf_link

- The second milestone deadline is **Thursday 20th of April, 2023 at 11:59 PM**.

Milestone 3

- In the third milestone, each team is requested to:
  1. Implement the ANN code for new data prediction and Classification.
  2. Tune the parameters (manually) till reaching the best results.
  3. Observe the effect of the parameters changing on the performance in terms of accuracy, MSE execution time, prediction time.

Milestone 3

- You are requested to submit:
  1. ANN Classification Python code,
  2. presentation **5 slides only** with the progress in this milestone (include graphs).

- All the files should be zipped and uploaded to the form:

https://docs.google.com/forms/d/e/1FAIpQLSeG50AZLDrPFSYc0byUqKE4KRn4nZlwXwxyt41uu7gpEMuOvg/viewform?usp=sf_link

- The third milestone deadline is **Thursday 27th of April, 2023 at 11:59 PM**.

Best of Luck! ☺