AI330_ Machine Learning Projects_Fall2023

Team Size:

Members: 6-7

Description:

As a team you will be implementing:

1. Linear regression & KNN as regressors on a numerical dataset.

2. Logistic regression & kmeans as classifiers on an image dataset (5 classes at maximum).

Grading:

Total: 20 marks

- **8 marks: Numerical dataset.**
- **8 marks: Image dataset.**
- **4 marks: Individual Assessment.**

Deliverables:

1) **Source Code + Datasets: Uploaded to GitHub, to be filled by the team leader later before discussion.**
2) **Project Cover Sheet:**
   It should include Faculty name, course name, team number, team members' IDs and names.
3) **Project Description Document:**
   For each model, you should specify:
   a. ***General Information on dataset:*** the name of dataset used, number of classes and their labels, the total number of samples in dataset and the size of each (in case of images), and finally the number of samples used in training, validation and testing.
   b. ***Implementation details:***
      - At feature extraction phase, how many features were extracted, their names, the dimension of resulted features.
      - Is cross-validation is used in any of implemented models? If yes, specify the number of fold and ratio of training/validation.
      - Hyperparameters used in your model, as initial learning rate, optimizer, regularization, batch size, no. of epochs, etc…

   c. Results details:
   For each model you should show all these results for your model on testing data (loss curve, accuracy, confusion matrix, ROC curve)

Datasets:

A) **numerical dataset of your selection (examples)**

B) **Choose image dataset from one of the following:**

1) **Cell Images for Detecting Malaria**
   a) **Link:** https://www.kaggle.com/datasets/iarunava/cell-images-for-detecting-malaria

2) **Columbia Object Image Library (COIL-100) Dataset**
   a) **Link:** http://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php

3) **Flower Species Recognition**
   a) **Link:** http://www.robots.ox.ac.uk/~vgg/data/flowers/102/index.html

4) **Traffic Sign Recognition**
   a) **Link:** http://people.ee.ethz.ch/~timofter/traffic_signs/index.html

5) **Fruits 360 Dataset**
   a) **Link :** https://www.kaggle.com/moltean/fruits

6) **Caltech-UCSD Birds-200 2011**
   a) **Link:** http://www.vision.caltech.edu/datasets/cub_200_2011/

7) **Character Recognition in Natural Images (The Chars74K dataset)**
   a) **Link:** http://www.ee.surrey.ac.uk/CVSSP/demos/chars74k/

8) **Oxford-IIIT Pet Dataset**
   a) **Link:** https://www.robots.ox.ac.uk/~vgg/data/pets/

9) **STL-10 dataset**
   a) **Link:** https://www.kaggle.com/jessicali9530/stl10

10) **Stanford Dogs Dataset**
   a) **Link:** http://vision.stanford.edu/aditya86/ImageNetDogs/main.html

11) **Age estimation**
   a) **Link:** https://susanqq.github.io/UTKFace/

12) **Plant Pathology 2020 – FGVC7**
   a) **Link:** https://www.kaggle.com/c/plant-pathology-2020-fgvc7/data

13) **Plant Disease Classification:**
   a) **Dataset:** PlantVillage Dataset

14) Food Recognition:
   a) **Dataset:** Food-101

15) Fine-Grained Bird Species Classification:
   a) **Dataset:** CUB-200-2011

16) Medical Image Diagnosis:
   a) **Dataset:** Diabetic Retinopathy Detection

17) Vehicle Make and Model Recognition:
   a) **Dataset:** Stanford Cars Dataset

18) Fashion Item Classification:
   a) **Dataset:** Fashion MNIST

19) Scene Classification:
   a) **Dataset:** MIT Scene Parsing Benchmark

20) Traffic Sign Recognition:
   a) **Dataset:** German Traffic Sign Recognition Benchmark

21) Mammogram Classification for Breast Cancer Detection:
   a) **Dataset:** Digital Database for Screening Mammography (DDSM)

22) Art Style Classification:
   a) **Dataset:** Painter by Numbers

23) Facial Expression Recognition:
   a) **Dataset:** Facial Expression Recognition Challenge (FER2013)

24) Tomato Detection:
   a) **Dataset:** *https://www.kaggle.com/datasets/nexuswho/tomato-detect*

25) Eye Diseases Classification
   a) **Dataset:** *https://www.kaggle.com/datasets/gunavenkatdoddi/eye-diseases-classification/data*