MLSL1 Assignment: Individual Assignment

This component has 50% weightage in the course.

Deliverables:

1. A **Jupyter NB (.ipynb)** with all the questions solved, **code executed**, and the relevant points written.

Note: The Assignment submission form should be submitted as well. Your submission will not be considered without the Assignment Submission form being submitted.

Instructions for the assignment:

1. [30 points] Visualizing Pairwise Fisher Discriminant
- **Dataset:** https://www.kaggle.com/c/digit-recognizer
- For every pair of classes (0, 1), (0, 9), …, (8, 9)
- Compute the Fisher Discriminant of each pixel (feature)
    - o Note that some of the pixels might have a zero denominator
    - o Treat that as 0 Fisher
- Normalize the remaining Fisher discriminant values from 0 to 1
- Draw the Fisher images of each of the pairs of classes
- See how they make sense (e.g., Class (0, 1), (3, 5), (4, 6)).
- This is "feature wise" Fisher goodness of each dimension

2. [25 points] Parameter Sweep over Decision Trees
- **Dataset:** https://www.kaggle.com/uciml/mushroom-classification
- Divide it randomly into 65% training and 35% test data
- Build a Decision Tree classifier with the following hyper-parameter combinations:
    - o Purity Measure: Entropy
    - o Maximum Depth:1, 2, 3, 4, 5, 6, 7
    - o Purity Threshold: 0.6, 0.7, 0.8, 0.9
    - o (Keep Size Threshold very low so only these two parameters are at play)
- Create a 2-D Table (Depth vs. Purity) above values
- In each cell, write the "Training" and "Test" accuracy
- Make an observation about the **best parameter combination**
- Draw the decision tree using the tools available
- Write down the rules learnt by this decision tree

3. [20 points] Parameter Sweep on k-NN
- **Dataset:** https://www.kaggle.com/c/digit-recognizer
- Divide it randomly into 65% training and 35% test data
- Build a k-NN classifier with the following hyper-parameters:
  - K = 1, 3, 5, 7, 9, 11, 13, 15
- Plot the training and test accuracies for each K
- Find the optimal K for this task

4. [25 points] Classification with PCA vs. FISHER using k-NN
- **Dataset:** https://www.kaggle.com/c/digit-recognizer
- Apply 9-dimensional PCA – Call it the Digits-PCA dataset.
- Apply 9-dimensional FISHER – Call it the Digits-FISHER dataset.
- Divide the data into 65% training and 35% test (after projection).
- Build a Decision Tree classifier with depth 5, purity threshold 0.8 for each dataset
  - What is the train and test accuracy of Digits-PCA-Tree classifier?
  - What is the train and test accuracy of Digits-FISHER-Tree classifier?
- Build a k-NN classifier with k = 5 on both datasets
  - What is the train and test accuracy of Digits-PCA-kNN classifier?
  - What is the train and test accuracy of Digits-FISHER-kNN classifier?
- Write Conclusions about projection methods to be used for classification

General Instructions:

1. This is an individual assignment.
2. Do NOT submit .zip files otherwise the submission will not be considered.
3. Late submission is applicable as per the course outline.
4. Please name your files properly and make sure your name and PGID is added to every submission made for this assignment.
5. The honour code for this submission is **2N-b**.
6. **Please look through the honor code restrictions carefully before attempting the assignment as they will be strong consequences for breaking them.**
7. **Please adhere to the given instructions, otherwise your submission will not be accepted, or severe penalty will be applied.**

Deadline: 19th February 2022, 11:55 pm.