

# MLSL1 Hands-on Assignment

## 1. [20points] 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. [20 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 (Im)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 Size Threshold: Very low
- 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:
  - o  $K = 1, 3, 5, 7, 9, 11, 13, 15$
- Plot the training and test accuracies for each K
- Find the optimal K for this task

## 4. [20 points] Classification with PCA vs. FISHER

- 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
  - o What is the train and test accuracy of **Digits-PCA-Tree** classifier?
  - o What is the train and test accuracy of **Digits-FISHER-Tree** classifier?
- Build a **k-NN classifier** with  $k = 5$  on both datasets
  - o What is the train and test accuracy of **Digits-PCA-kNN** classifier?
  - o What is the train and test accuracy of **Digits-FISHER-kNN** classifier?
- Build a **Logistic Regression** with between any pair of classes (say 0 vs 3, 4 vs 9)
  - o What is the train and test accuracy of Digits-PCA-Logistic-Regression classifier?
  - o What is the train and test accuracy of Digits-FISHER-Logistic-Regression classifier?
- Comment on whether PCA vs. FISHER is better for projection