- **Target:** Implement (semi-)parametric and non-parametric classification.
- **Dataset:** Download points2d.dat from `https://www.cmpe.boun.edu.tr/~emre/courses/cmpe462/material/hws/hw2/points2d.dat`. The sample includes 6000 instances. Each instance is represented by a feature vector of 2 dimensions. The third column corresponds to the class (0/1/2) of the instance.
- **Partitioning:** Split data into training (50%), validation (25%), and test (25%) sets. Use training and validation sets for finding the best model and test set for providing the prediction error.
- **Methods:**
  1. (4 pt) Implement mixture of Gaussians. Run it with k=1,2,3 (number of Gaussians - might differ for different classes) using different initial points (the general practice is to use k-means to initialize the means). Report (for the best model):
     - the number of Gaussians for class 0, 1 and 2,
     - the prediction error, and
     - the confusion matrix.
  2. (4 pt) Implement k-nn. Run it with k=1,10,40. Report
     - best k,
     - prediction error for all k's.
     - confusion matrix for all k's.
  3. (0.5pt) First submission / mistake report bonus
- **Submission:**
  - Create an archieve named **student-id.tgz** with the contents (below) placed in a directory named **student-id/**. Send the archieve file through email (subject:**hw2 submission student-id**) to the instructor.
  - The source code should be a single file named **student-id.x**. Feel free to use matlab/octave/python/R. The source code should be properly commented. Use of built-in functions are limited to the very basic ones. You need to implement mixture of Gaussians and k-nn algorithms yourself.
  - The report should be named as **student-id.pdf**. The report for each method should include (i) a paragraph that explains the implementation details (e.g. details of multi-class classification, initialization, etc.), and (ii) the answers to the questions above (prediction error, etc.).