

## Project 3 - Classification Using Non-Parametric Density Estimation and Performance Evaluation - Due 03/20/18

### Basic requirement (80)

- Task 1 (20 pts): Implement kNN
  - The implementation should be able to flexibly change the value of "k"
  - The implementation should be able to measure the run time
- Task 2 (20 pts): Experimenting kNN using the pima dataset
  - Apply kNN on nX, pX, and fX (refer to project 2) and compare the performance with that from project 2 from both computational time and accuracy perspectives
  - For each application, try different values of k, e.g.,  $k=1$ , 5, and  $\sqrt{n}$ . Use Euclidean and other Minkowski distance of different degrees (Refer to Lecture note 10).
- Task 3 (20 pts): Use the "fglass" data set. Experimenting cross-validation with kNN as the classifier. Try out different k's. Based on the performance from cross-validation, determine the best k for this data set. You need to get enough samples of k between 1 and  $\sqrt{n}$ , e.g., 1, 5...15, and  $\sqrt{n}$ .
- Task 4 (20 pts): Implement NB classifier fusion approach and report the best fusion result.

### Report (20)

### Bonus (+15)

- Design and implement methods other than discussed in class to reduce the computational load of kNN. Show the time spent and the storage used in the new implementation. Also compare the accuracy.