ECE 662, 2020S, Mini Project 2

1. Task 1

Data generation method: Consider N-dimensional feature vectors coming from C classes. Assume that the distributions of the feature vectors for the two classes are (known) normal distributions (with same priors).

Study Parzen-window method:

- C=2, 80 samples for each class (40 for training, 40 for testing): Evaluate and plot error rate (@ testing data) for varying Parzen window sizes (consider at least 4 different dimension : N)
- C=2, 1000 samples for each class (40 for training, 40 for testing): Evaluate and plot error rate (@ testing data) for varying Parzen window sizes (consider at least 4 different dimensions : N)
- C=5, N =2, 80 samples for each class (40 for training, 40 for testing): plot error rate (@ testing data) for varying Parzen window sizes to analyze the best window size
- Analyze how # of training samples (e.g., from 10 to 10k) impact the error rate (@ testing data), with different dimension: N. You can choose other parameters based on your own need.

2. Task 2

You can use the method in Task 1 for data generation, for reuse its data.

Study K-NN method:

- C=2, 80 samples for each class (40 for training, 40 for testing): Evaluate and plot error rate (@ testing data) for varying K (consider at least 4 different dimension : N)
- C=2, 1000 samples for each class (40 for training, 40 for testing): Evaluate and plot error rate (@ testing data) for varying K (consider at least 4 different dimensions : N)
- C=5, N =2, 80 samples for each class (40 for training, 40 for testing): plot error rate (@ testing data) for varying K to analyze the best K
- Analyze how # of training samples (e.g., from 10 to 10k) impact the error rate (@ testing data), with different dimension: N. You can choose other parameters based on your own need.
- Study the difference of Euclidean distance and Manhattan distance (https://en.wikipedia.org/wiki/Taxicab_geometry), you can choose C=2, varying K. You can choose other parameters based on your own need.

Notes:

- Due date March 30th, 11:59pm
- Group work (max 3 persons) is encouraged, but separate report needs to be submitted (with different parameters). Also add your team members' name
- Electronic submission to canvas
- Submission: a project report (doc, docx, or PDF), all executable code files, a readme file to introduce the code
- Matlab, or python is suggested. Other languages also acceptable.
- All codes need to have detailed comments in the file, especially close to your parameters, and subfunctions
- A readme file to introduce the codes files are required, explaining the flow and dependency