CMPE 544 Pattern Recognition
Spring 2025 Assignment 2
Due May 18th by midnight

Classification

In this task, you will again use the Quick Draw dataset given to you in the first assignment. The dataset is still available at `https://drive.google.com/file/d/1oG0HnQEb7VVwpdYsn8rShfAQ8Vq4hA view?usp=sharing`. You may use the same features you extracted in Assignment 1, or explore better features if necessary.

1. (25 pts) In this question, please focus only on rabbit and hand classes. Please implement a linear soft-margin SVM from scratch using a quadratic programming solver.
   (a) You may use any library you want for the quadratic programming solver. Please clearly write the expressions you feed to the solver.
   (b) Please tune the parameter C using 5-fold cross-validation.
   (c) Please report the binary classification test accuracy and the running time of the training.
2. (20 pts) Now, you may use Scikit-learn's SVM function again for rabbit and hand classes. Please train a linear SVM and various SVMs with non-linear transformations.
   (a) Please tune the hyperparameters using 5-fold cross-validation.
   (b) Compare the binary classification test accuracy and the running time of the training with the first part.
3. (15 pts) In this question, you may use the Scikit-learn's SVM function in the second question. Please take the best-performing classifier in the second question and obtain the support vectors. Inspect the images that are the support vectors and comment on the visual differences between support vectors and samples far away from the margin.

Clustering

In this task, you will use the same dataset and features with all of its five classes.

1. (20 pts) Please implement the k-means clustering algorithm and cluster the dataset with $k = 5$. Please use Euclidean distance as the distance metric.
2. (10 pts) Report the clustering performance with the sum of squared error (SSE) and clustering accuracy. Please search the literature for two more internal and two more external metrics in addition to SSE and clustering accuracy.
3. (10 pts) Please try different distance metrics, such as Manhattan and Cosine distance metrics. Report their performances using the evaluation metrics in the second question. Please compare different distance metrics in terms of performance and how they measure the distance between two data points.

Submission

- A PDF report addressing the above points, including your name, student number, references, and drive link to the code, should be submitted.
- You should explain how you implemented the classifiers and solved issues in your report.
- Please include the link to your code on the first page of your report.
- Your experiments should be reproducible.
- A readme should be provided.

IMPORTANT NOTES:

- There is no report template. However, your reports should address all the bullet points above.
- You should not forget to cite your references. If you followed a GitHub repository, do not forget to cite it.
- Please note that you should be ready to answer questions regarding your code.