

# Face and Digit Classification

Final Project

**Michael Li - 192008938**

**Daniel Liu - 184007283**

**Robert Kulesa - 185009892**

CS440 Fall 2021

Professor Boularias

Rutgers University - New Brunswick

December 14, 2021

## Classifier 1 - Perceptron

## Classifier 2 - Naive Bayes

## Classifier 3 - K Nearest Neighbor

The K-Nearest Neighbor (KNN) Classifier works by taking a test sample, calculating its distance to all training samples, and classifying the test sample as whichever label the majority of the  $K$  closest samples hold.

{include section on feature extraction if we all do it differently}

Many distance functions can be used, and each has their own advantages and disadvantages for different datasets, such as speed and memory requirements.

For the face dataset, this implementation of KNN uses the cosine distance as the distance function. Mathematically, the cosine distance is the cosine of the angle between two vectors of length  $n$  in  $n$ -dimensional space. The cosine distance of vectors  $q$  and  $p$  can be calculated as:

$$\text{dist}(q, p) = 1 - \cos(\theta) = 1 - \frac{q \cdot p}{\|q\| \|p\|}$$

For the digits dataset, this implementation of KNN uses the euclidean distance as the distance function. Mathematically, the euclidean distance is the length of the line segment that connects two points. The euclidean distance of vectors  $q$  and  $p$  can be calculated as:

$$\text{dist}(q, p) = \sqrt{\sum_{i=1}^n (q_i - p_i)^2}$$

## Classification Model Comparison