Machine Learning Project 1, KNN

In this project you will practice the basics of Machine Learning Classification by creating a K-NN classifier for two datasets. You will also practice good practices for how to describe, evaluate, and write up a report on the classifier performance.

Datasets: The project will explore two datasets, the famous MNIST dataset of very small pictures of handwritten numbers, and a dataset that explores the prevelance of diabetes in a native american tribe named the Pima. You can access the datasets here:

1. `https://www.kaggle.com/uciml/pima-indians-diabetes-database`
2. `https://www.kaggle.com/c/digit-recognizer/data`

Task: For each dataset, you must create a K-NN classifier that uses the training data to build a classifier, and evaluate and report on the classifier performance. You should turn in a report with the following sections.

Report Sections

Header: List your name, the major resources you used to complete this project, the programming language you used.

Dataset details: Describe the data and some simple visualizations (for images, a few examples from each category; for other data, perhaps some scatter plots or histograms that show a big picture of the data). Describe your training, validation and testing split and justify your choices.

Algorithm Description: K-NN is a very clear algorithm, so here describe any data pre-processing, feature scaling, or otherwise that you did.

Algorithm Results: Show the accuracy of your algorithm — in the case of the Pima Dataset, show accuracy with tables showing false positive, false negative, true positive and true negatives. In the case of the MNIST digits show the complete confusion matrix. Choose a single number to measure accuracy and show how that number varies as a function of K.

Runtime: Describe the run-time of your algorithm and share the actual "wall-clock" time that it took to compute your results.

It is expected that your project report may require 2 pages per dataset if you are good about making interesting figures and making them not too large, or 3-4 pages if your figures are big. The LaTex that generated this page is available here: `https://v2.overleaf.com/read/wxkggcjxjgtw`. This is a living document, and we may add to it over time.