**DENSITY ESTIMATION AND CLASSIFICATION**

Francis Mendoza

ASUID: 1213055998

CSE575- Statistical Machine Learning

fmendoz7@asu.edu

1. **INTRODUCTION**
   1. I extracted a subset of the MNIST dataset, particularly dealing with the digits “7” and “8”. In this project, I used MLE Density Estimation, Naïve Bayes classification and Logistic Regression to find the average of all pixel values in the image, as well as the standard deviation of all pixel values in the image.
2. **FEATURE EXTRACTION**
   1. The extracted dataset contains training and testing samples for “7” and “8”, stored as a dictionary.
      1. The key is a string literal, denoting training or testing and which digit
      2. The value is a list of lists, denoting the brightness output of pixels
   2. Training Samples:
      1. “7”: 6265 rows
      2. “8”: 5851 rows
   3. Test Samples:
      1. “7”: 1028
      2. “8”: 974
   4. The results from the initial feature extraction (calculating the means of both the means and the standard deviations) follows as such:

**TRAINING SET**

* + 1. 0.3206804173181901
* Since the covariance is small, this implies that there is no strong correlation between mean and standard deviation (and thus, are strongly independent variables)

**TESTING SET**

1. 0.3238037438846891
2. **NAÏVE BAYES CLASSIFICATION**
   1. \_
3. **LOGISTIC REGRESSION**
   1. \_