In the name of God

Shiraz University

Pattren Recognition

#Mini Project 2

Maximum-Likelihood & Bayesian Parameter Estimation

Pattern.recognition.1399@gmail.com

Due Date 24 th Azar

Problem #1

1.1. 1.Design and implement a Bayes optimal classifier with a Gaussian parametric estimate of pdfs to minimize the probability of classification error using a reduced TinyMNIST dataset. You must state the equations which are used for the parameter estimation, and also explain how you choose the prior probabilities of the classes.

1.2. When estimating a Gaussian distribution parameters, sometimes a singular matrix is obtained as the covariance of the data.

a)Why this situation is problematic?

b)This difficulty arises for the given dataset. Using the following hint, study the proposed methods, and apply one of them to your classifier. Evaluate your classifier.

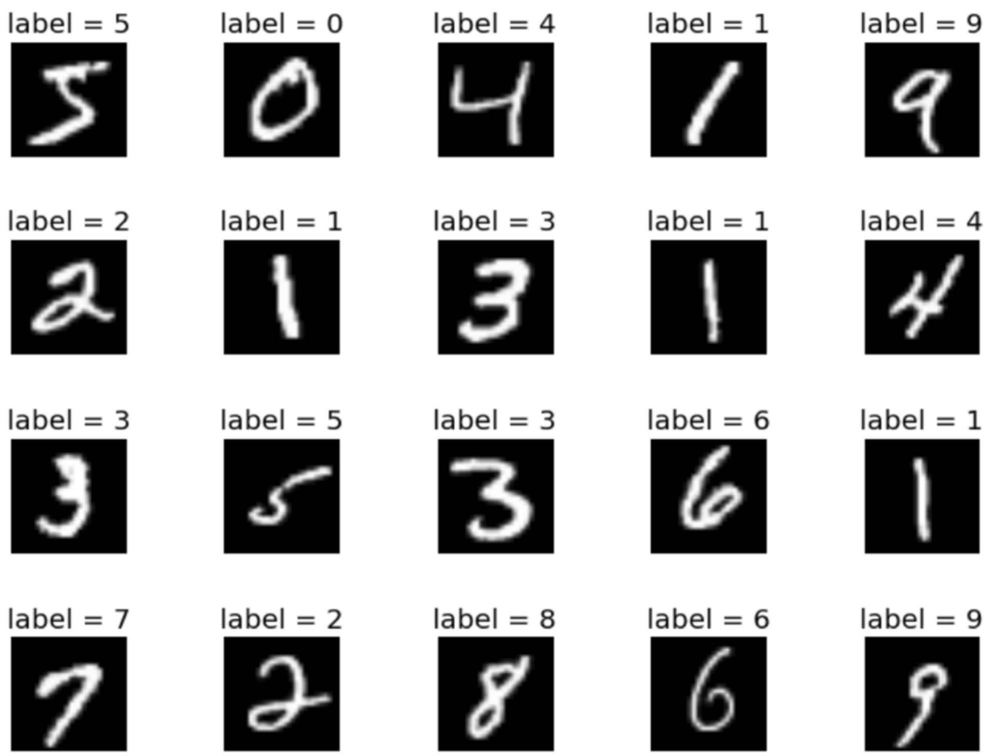
Hint:
https://www.doc.ic.ac.uk/~dfg/ProbabilisticInference/old_IDAPILecture16.pdf

Dataset Description

The **MNIST** database of handwritten digits (including digits 0 to 9, so has ten classes) has a training set of 60,000 examples and a test set of 10,000 examples. It is a subset of a larger set available from NIST. The digits have been size normalized and centered in a fixed-size image (28*28).

It is a good database for people who want to try learning techniques and pattern recognition methods on real-world data while spending minimal efforts on preprocessing and formatting.

The original dataset, named “MNIST,” and the small version, we name “**TinyMNIST**”. The TinyMNIST dataset contains 5,000 train samples and 2,500 test samples, and the images are scaled from 28*28 to 14*14 pixels, which results in 196 features.



- For the sake of simplicity, 62 features are extracted from the TinyMNIST Datset. You only need to work with thease extracted features in your assignments.

- The dataset folder contains 4 “.csv” files, “test_Data.csv”, “test_Labels.csv”, “train_Data.csv”, and “train_Labels.csv”. The “Data” files are matrices of size (Number of Samples, Number of Features). The “Labels” files are vectors of size (Number of Samples) containing label for each sample.

- The dataset is already shuffled and normalized to the range (0, 1), so there is no need to reshuffle or normalize it.

- In order to estimate parameters of the Gaussin distribution, you may use training data. You should also use test data for evaluating your classifier.

Problem #2

Apply MAP estimation to calculate $p(\theta|D)$ and $p(X|D)$ Under the assumption that $p(X|\theta)$ follows a normal distribution $p(X|\theta) = p(X|\mu) \sim N(\mu, \sigma^2)$ and $p(\mu) \sim N(\mu_0, \sigma_0^2)$ and Proof the equation below:

Data : a set D of N (i.i.d) training samples $X_1, X_2, \dots, X_N$.

$$p(\mu|D) = \frac{p(D|\mu)p(\mu)}{p(D)} = \prod_{k=1}^{N} p(X_k|\mu)p(\mu) = \alpha \, \frac{1}{\sqrt{2\pi\sigma_N^2}} e^{[-\frac{1}{2}\frac{(\mu-\mu_N)^2}{\sigma_N^2}]}$$

$* \ \bar{\mu} = \frac{1}{N}\sum_{k=1}^{N} X_k$

$* \ \mu_N = \frac{N\sigma_0^2}{N\sigma_0^2+\sigma^2}\bar{\mu} + \frac{\sigma^2}{N\sigma_0^2+\sigma^2}\mu_0$

$* \ \sigma_N^2 = \frac{\sigma^2\sigma_0^2}{N\sigma_0^2+\sigma^2}$
