

## PRML Midterm (Experiments & Report )

(Due on Saturday, June 6, 2020.)

Please submit to our TA—Siwen Liu before 8:30 am.

For every 5 minutes beyond the deadline, 5 points will be deducted from your score.

### Problem 1: Maximizing-A-Posterior (MAP) Decision Rule and Maximum-likelihood Estimation (MLE) (35 points)

In Chapter 3, we learnt how to design an optimal classifier based on MAP decision rule if we knew the prior probabilities and the class-conditional densities of all categories (denoted as  $\omega_i, i = 1, 2, \dots, c$ ). However, in practical pattern recognition applications we rarely have this kind of complete knowledge. Therefore, we introduced parameter estimation approaches which illustrated how to use data samples to estimate the unknown probabilities and probability densities.

Let us consider a three-category classification problem, and assume that the three categories have equal priors. The sampled data are stored in three files, data1.txt, data2.txt, and data3.txt (Note: the files can be downloaded from the shared folder of the QQ group, midterm.zip). The first file contains the samples of class 1, the second file contains the samples of class 2 and so forth. Each sample has two dimensions and each file contains 2000 samples. In the following exercises, we will use part of the samples to train three MAP classifiers, one for each category, and then evaluate the classifiers with the rest of the samples.

- (a) (0 points) Warming up. Plot the first 1000 samples of each category. Your result should be similar to Figure 2.
- (b) (10 points) Since  $P(\omega_1) = P(\omega_2) = P(\omega_3)$ , the posterior probabilities  $P(\omega_i|\mathbf{x}), i = 1, 2, 3$ , are determined by the class-conditional densities  $p(\mathbf{x}|\omega_i), i = 1, 2, 3$ . Hence, we first estimate/train  $p(\mathbf{x}|\omega_i), i = 1, 2, 3$  with MLE approach. Assume that

$$p(\mathbf{x}|\omega_i) \sim N(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i).$$

Use the first 1000 samples of each category to estimate  $p(\mathbf{x}|\omega_i)$  and plot the resulting class-conditional densities.

- (c) (10 points) Use MAP decision rule to classify the rest of the samples and show the rate of misclassification in each category.
- (d) (10 points) If we use the first 500 samples of each category for training, and the rest for testing. Repeat (b) and (c), and show the rate of misclassification in each category.

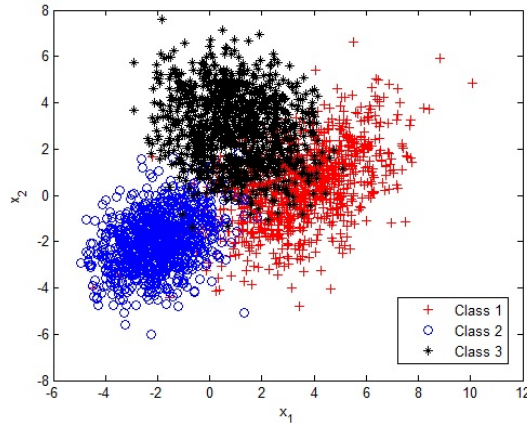


Figure 1: The first 1000 samples of each category.

- (e) (5 points) Describe and interpret your findings by comparing the results of (c) and (d).

**Problem 2: Parzen Window Estimation and  $k$ -Nearest Neighbor ( $k$ -NN) Estimation (40 points)**

Let us revisit the three-category classification in Problem 1, but utilize non-parametric methods to perform classification this time. We assume that the three categories have equal priors. The sampled data are stored in three files, data1.txt, data2.txt, and data3.txt (Note: the files can be downloaded from the shared folder of the QQ group, midterm.zip). The first file contains the samples of class 1, the second file contains the samples of class 2 and so forth. Each sample has two dimensions and each file contains 2000 samples.

- (a) (10 points) In Chapter 5, one important conclusion about parzen window estimation method is “ $p_n(\mathbf{x})$  is actually the average of  $n$  normalized window functions, each centered at  $\mathbf{x}_i$ ” where  $p_n(\mathbf{x})$  denotes the estimate of the true density function and  $\mathbf{x}_i$  denotes the  $i$ -th sample ( $1 \leq i \leq n$ ). Alternatively, the parzen window estimate  $p_n(\mathbf{x})$  can be expressed as:

$$p_n(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^n \delta_n(\mathbf{x} - \mathbf{x}_i),$$

where  $\delta_n(\cdot)$  denotes the normalized window function. Suppose the window function is a multivariate Gaussian function,

$$\delta_n(\mathbf{x}) = \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^t \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right],$$

where  $\boldsymbol{\mu} = \mathbf{0}$  and  $\boldsymbol{\Sigma} = \mathbf{I}$  ( $\mathbf{I}$  is the identity matrix). Use the first 1000 samples of each category to obtain  $p_n(\mathbf{x})$  for each category and plot the parzen window estimate of the density function of each category.

- (b) (10 points) Use the results of (a) and Maximizing-A-Posterior (MAP) decision rule to classify the rest of the samples and show the misclassification rate of each category.
- (c) (10 points) In Chapter 5, we also introduced how to use  $k$ -NN method to estimate the posterior distribution. Suppose we are given the first 1000 samples of each category as training samples, and asked to classify the rest of the samples using NN classifier (i.e.  $k=1$ ). Show the misclassification rate of each category.
- (d) (5 points) Repeat (c) when  $k = 10$ . (Hint:  $P(\omega_i|\mathbf{x}) = k_{\omega_i}/k$ , where  $k_{\omega_i}$  is the number of samples labeled  $\omega_i$  within the  $k$  nearest neighbors of  $\mathbf{x}$ )
- (e) (5 points) Describe and interpret your findings by comparing the results of (c) and (d).

### Submission Guidelines

1. Create a zip file, YourID\_midterm.zip, which includes your source code and a short report that clearly illustrates the required results. The report has to be in the format of pdf or doc. In addition, please verify that all the files can be successfully extracted from the zip file before submission.