ENGG 5202: Homework #2

Due on Thursday, March 10, 2016, 5:30pm

Problem 1

[**30 points**]

Consider a one-dimensional two-category classification problem with priors, $P(\omega_1) = 1/2$ and $P(\omega_2) = 1/2$. The class conditional densities have the form

$$p(x|\omega_i) = \begin{cases} 0 & x < 0 \\ \theta_i e^{-\theta_i x} & x \geq 0 \end{cases}$$

1. We have training data $D_1 = \{2, 4\}$ for class ω_1 and $D_2 = \{5, 7\}$ for class ω_2. Find the maximum-likelihood estimates $\hat{\theta}_1$ and $\hat{\theta}_2$ for θ_1 and θ_2.
2. Given the parameters of the maximum-likelihood estimation, find the decision boundary x^* for the minimum classification error.
3. What's the expected classification error of your classifier in question 2?

Problem 2

[**10 points**]

$\mathbf{A} = [a_{ij}]$ is the transition matrix of HMM. When using EM to estimate the parameters of HMM, a_{12} is initialized as 0. Prove that a_{12} will remain zero in all subsequent updates of the EM algorithm.

Problem 3

[**30 points**]

To understand the "curse of dimensionality" in greater depth, consider the effects of high dimensions on the Parzen window density estimation. Suppose we need to estimate a density function $p(x)$ in the unit hypercube in $\mathcal{R}^d$ based on n samples.

1. Let n_1 denote the number of samples in a "dense" sample in $\mathcal{R}^1$. What is the sample size for the "same density" in $\mathcal{R}^d$? If $n_1 = 100$, what sample size is needed in a 20-dimensional space?
2. Consider points uniformly distributed in the unit hypercube. Find $l_d(p)$, the length of a hypercube edge in d dimensions that contains the fraction p of points ($0 \leq p \leq 1$). To better appreciate the implications of your result, calculate: $l_5(0.01)$, $l_5(0.1)$, $l_{20}(0.01)$, and $l_{20}(0.1)$.
3. Based on the observations above, discuss how high dimensionality would affect the Parzen window density estimation.

Problem 4

[**30 points**]

Fisher linear discriminant (FLD) can be used to find a linear discriminant function $g(\mathbf{x}) = \mathbf{w} \cdot \mathbf{x} + w_0$. $\mathbf{w}$ is found by FLD. After projecting all the samples to a one-dimensional space, the threshold w_0 can be found by multiple possible ways. In this problem we will compare FLD with logistic regression, both of which are used to train linear classifiers. You are provided with some matlab data and functions to do experimental evaluation.

The "data.mat" file contains two data sets. Each data set contains a training set (**train*i***) and a test set (**test*i***). **train1**, **train1_2** and **test1** are generated from two 2D Gaussians with the same covariance. **train1_2** is identical to **train1** except that it has an extra training point not sampled from the Gaussian. **train2** and **test2** are from a data set for digit classification. A digit image is represented by a 64-dimensional vector of 1's or 0's - pixels in a 8×8 bitmap. For all the training and testing data, the .X field of each variable is the representation of data samples and the .y field contains the class labels (0 or 1).

The descriptions of matlab functions are as following.

- **plotdata(train*i*)** can plot 2D data and associated labels.
- **w = fisherdiscriminant(train*i*.X, train*i*.y)** trains a FLD classifier and returns its parameters. $\mathbf{w}(1)$ is the threshold w_0 of the linear classifier.
- **w = logisticreg(train*i*.X, train*i*.y)** trains a logistic regression classifier.
- **boundary([w1 w2 ...], test*i*)** plot the 2D data and the decision boundary of several linear classifiers in one figure.
- **errorrate(w, test*i*)** computes the error rate of a linear classifier.

1. Suppose that there are enough training samples. With certain assumptions on the underlying distributions of classes, the decision boundary found by FLD is the same as that found by the Bayesian decision theory. What are the assumptions?

Problem 4 continued on next page...

2. Train FLD and logistic regression on **train1** and **train1_2** and compute the error rates on **test1** using the obtained four classifiers. Plot the decision boundaries and the training data. What do you observe from the experimental result? Do FLD and logistic regression have similar testing error rates when being trained using **train1**? Explain why. Do their error rates change significantly when being trained on **train1_2**? Explain why.

3. Train FLD and logistic regression on **train2** and compute their error rates on **test2**. Compare the error rates of the two classifiers and explain why they behave similarly or differently on this data set. Is the Gaussian assumption reasonable for this representation of digit images? Why? Are FLD and logistic regression sensitive to the Gaussian assumption? Why?