CSE 555 Spring 2010 Mid-Term Exam

Jason J. Corso
Computer Science and Engineering
SUNY at Buffalo
jcorso@buffalo.edu
Date 5 Mar 2010

The exam is worth 100 points total and each question is marked with its portion. The exam is closed book/notes. You have 50 minutes to complete the exam. Use the provided white paper, write your name on the top of each sheet and number them. Write legibly.

Problem 1: "Recall" Questions (25pts)

Answer each in one or two sentences **max**.

1. (5pts) Define decision boundary.
2. (5pts) In Bayes Decision Theory, what does the prior probability capture?
3. (5pts) What is the zero-one loss function?
4. (5pts) What is the basic idea of the curse of dimensionality?
5. (5pts) What does LLE focus on preserving during learning?

Problem 2: Parameter Estimation (15pts)

Parameter estimation is a key aspect of pattern classification.

1. (3pts) In parameter estimation, what is the difference between Maximum A Posteriori and Bayesian parameter estimation?
2. (6pts) For a dataset $\mathcal{D}$ of n samples, the goal of Bayesian parameter estimation is to estimate $p(\mathbf{x}|\mathcal{D})$ for any x. We go about this by marginalizing over the parameters:

$$p(\mathbf{x}|\mathcal{D}) = \int p(\mathbf{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\mathcal{D})d\boldsymbol{\theta}. \tag{1}$$

   Explain the various terms of this equation. What are the assumptions required to make this work?

3. (6pts) Recursive Bayesian methods. Consider a univariate problem in which we assume our samples come from a uniform distribution:

$$p(x|\theta) \sim U(0,\theta) = \begin{cases} 1/\theta & 0 \le x \le \theta \\ 0 & \text{otherwise} \end{cases}. \tag{2}$$

Initially, assume it is a non-informative prior bounded at the top at 10, i.e., $0 < \theta \leq 10$. Use recursive Bayes to estimate θ for the following two samples of data $\mathcal{D} = \{5, 2\}$. Give the analytical form after each iteration of the recursion.

Problem 3: Discriminant Functions (40pts)

This problem is about discriminant functions.

1. Consider the general linear discriminant function

$$g(x) = \sum_{i=1}^{\hat{d}} a_i \phi_i(\mathbf{x}) \tag{3}$$

with augmented weight vector $\mathbf{a}$. Let $\mathbf{y} = [\phi_1(\mathbf{x}) \dots \phi_{\hat{d}}(\mathbf{x})]^\mathsf{T}$.

   (a) (3pts) What are the role of the ϕ functions?

   (b) (3pts) Write the equation for the plane in $\mathbf{y}$-space that separates it into two decision regions.

   (c) (3pts) A weight vector $\mathbf{a}$ is said to be a solution vector if $\mathbf{a}^\mathsf{T}\mathbf{y}_j > 0 \quad \forall j \in 1, \dots, n$. In general, is this solution vector unique? Why?

2. Consider the relaxation criterion function, let b be a margin,

$$J_r(\mathbf{a}) = \frac{1}{2} \sum_{\mathbf{y} \in \mathcal{Y}} \frac{\left(-\mathbf{a}^\mathsf{T}\mathbf{y} - b\right)^2}{\|\mathbf{y}\|^2} \tag{4}$$

   (a) (3pts) Compare this relaxation criterion function to the perceptron criterion function.

   (b) (4pts) For a single-sample relaxation learning procedure, what is the update rule?

   (c) (4pts) What is the geometrical interpretation of the update rule? Draw a figure to help explain.

3. (20pts) Relating the Minimum-Squared Error procedure for training generalized linear discriminants to the Fisher linear discriminant. Recall the MSE procedure for estimating a discriminant. In matrix Y let each row be a data sample, and let $\mathbf{b}$ be a vector of margin values for each point. The MSE procedure turns the inequalities into equalities:

$$Y\mathbf{a} = \mathbf{b} \tag{5}$$

Ultimately, the solution we seek minimizes the sum-of-squared error criterion over all of the samples:

$$J_s(\mathbf{a}) = \sum_{i=1}^{n} (\mathbf{a}^\mathsf{T}\mathbf{y}_i - b_i)^2. \tag{6}$$

Taking the derivative and equating it to 0 gives us the necessary conditions:

$$Y^{\mathsf{T}} Y \mathbf{a} = Y^{\mathsf{T}} b. \tag{7}$$

The pseudoinverse is how we solve it.

$$\mathbf{a} = (Y^{\mathsf{T}} Y)^{-1} Y^{\mathsf{T}} \mathbf{b}. \tag{8}$$

Consider the following specific augmentation and selection of the margin vector. Assume (1) the augmentation is to simply add a constant 1 to the top of each sample vector, (2) we normalize the class ω_2 samples by multiplying by -1, and (3) the first n_1 samples are labeled ω_1 and the second n_2 samples are labeled ω_2. X_1 is the matrix of class 1 samples x with each row a sample, and X_2 is the matrix for class 2 samples. $\mathbf{1}_i$ is a column vetor of n_i ones.

$$Y = \begin{bmatrix} \mathbf{1}_1 & X_1 \\ \mathbf{1}_2 & X_2 \end{bmatrix} \tag{9}$$

Let $\mathbf{b}$ be set as the following.

$$\mathbf{b} = \begin{bmatrix} \frac{n}{n_1} \mathbf{1}_1 \\ \frac{n}{n_2} \mathbf{1}_2 \end{bmatrix} \tag{10}$$

Show that for this choice of the margin vector, the MSE solution is equivalent, up to a scale factor, to the Fisher linear discriminant.

Problem 4: General Pattern Recognition (20pts)

For the two-class problem (cross and solid) in the figure below, describe how you might design a classifier based on ideas we've covered.