CS 274A Homework 4


Due Date: Monday February 29th (at the start of class)

Guidelines for Homeworks

- The homeworks are intended to help you work through the concepts we discuss in class in more detail. It is important that you try to solve the problems yourself. The homework problems are important to help you better learn and reinforce the material from class. If you don’t do the homeworks you will likely have difficulty in the exams later in the quarter.

- If you can’t solve a problem, you can discuss it verbally with another student. However, please note that before you submit your homework solutions you are not allowed to view (or show to any other student) any written material directly related to the homeworks, including other students’ solutions or drafts of solutions, solutions from previous versions of this class, and so forth. The work you hand in should be your own original work.

- You are allowed to use reference materials in your solutions, such as class notes, textbooks, other reference material (e.g., from the Web), or solutions to other problems in the homework. It is strongly recommended that you first try to solve the problem yourself, without resorting to looking up solutions elsewhere. If you base your solution on any material that we did not discuss in class, or is not in the class notes, then you need to provide a reference, e.g., “based on material in Section 2.2 in .....”

- In problems that ask for a proof you should submit a complete mathematical proof (i.e., each line must follow logically from the preceding one, without “hand-waving”). Be as clear as possible in explaining your notation and in stating your reasoning as you go from line to line.

- If you wish to use LaTeX to write up your solutions you may find it useful to use the .tex file for this homework that is posted on the Web page.

- Please answer all of the questions and submit a hardcopy of your written solutions at the beginning of class (either hand-written or typed are fine as long as the writing is legible). Clearly mark your name on the first page. Code (if requested) should be submitted to the EEE dropbox.

- All problems are worth 10 points unless otherwise stated. All homeworks will get equal weight in computation of the final grade for the class.
Problem 1:

In practice regularization is often used in regression problems to prevent the weights from getting too large, e.g., minimize the empirical loss function plus $\lambda \sum_{j=1}^{d} \theta_j^2$, where $\lambda$ (the amount of regularization) is determined via cross-validation. For minimizing the log-loss function with logistic regression (and $y \in \{0, 1\}$) derive the gradient update equation (batch, not stochastic) for each weight $\theta_j$ when regularization in the form $\sum_{j=1}^{d} \theta_j^2$ is being used with some fixed $\lambda$ value.

Problem 2:

Consider a 2-class classification problem with $d$-dimensional real-valued inputs $x$, where the class-conditional densities, $p(x|c_1)$ and $p(x|c_2)$ are multivariate Gaussian with different means $\mu_1$ and $\mu_2$ and a common covariance matrix $\Sigma$, with class probabilities $P(c_1)$ and $P(c_2)$.

1. Write the optimal discriminant function $g_1(x)$ in the form $g_1(x) = \log p(x|c_1) + \log p(c_1)$ where the right-hand side is expressed in terms of the parameters of the model.

2. Prove that the optimal decision boundary, at $g(x) = g_1(x) - g_2(x) = 0$, can be written in the form of a linear discriminant, $w^T x + w_0 = 0$, where $w$ is a $d$-dimensional weight vector and $w_0$ is a scalar, and clearly define $w$ and $w_0$ in terms of the parameters of the classification model.

3. Show that if $P(c_1) = P(c_2)$ then the classifier can be viewed as a “nearest mean” type of classifier, where a vector $x$ is assigned to the mean that it is closest to, and where distance is measured using Mahalanobis distance.

Problem 3:

Using the same notation and problem definition as in the previous problem (2-class problem, Gaussians with a common covariance matrix, arbitrary mean vectors), prove that the logistic regression function, as a function of the vector $x$, is an optimal discriminant function for this problem (when the weights in the logistic function are defined appropriately). You can use the results from the previous problem in your solution here.

Problem 4:

Consider a 2-class classification problem with 2-dimensional real-valued inputs $x$, where the class-conditional densities, $p(x|c_1)$ and $p(x|c_2)$ are multivariate Gaussian with class probabilities $P(c_1)$ and $P(c_2)$. Consider the case where the mean of class 1 is at $(1, 1)$ and the mean of the second class is at $(4, 4)$, and with a covariance matrix defined so that $\sigma_{11} = 1$, $\sigma_{22} = 4$, and $\sigma_{12} = 1$. Draw or plot a figure in the 2-dimensional space $(x_1, x_2)$ that shows
1. the location of the optimal decision boundary relative to the two class means, with \( p(c_1) = p(c_2) \).

2. the location of the optimal decision boundary, with \( p(c_1) = 0.8, p(c_2) = 0.2 \). Provide a 1-line interpretation of why the result is different to part 1.

If you wish you can use MATLAB to generate the plots for you, but hand-drawn is also fine.

**Problem 5:**

Consider a classification problem with 2 classes \( c_1 \) and \( c_2 \) and a single real-valued feature vector \( x \), where class 1 has a Gaussian density \( p(x|c_1) \) with parameters \( \mu_1 \) and \( \sigma^2_1 \), and class 2 has a Gaussian density \( p(x|c_2) \) with parameters \( \mu_2 \) and \( \sigma^2_2 \).

1. Derive a general expression for the location of the optimal decision regions as a function of the parameters.

2. Now assume \( \mu_1 = 0 \) and \( \sigma^2_1 = 1 \) and \( \mu_2 = 3 \) and \( \sigma^2_2 = 3 \). Also assume \( p(c_1) = p(c_2) = 0.5 \). Draw or plot each of \( p(x|c_1)p(c_1) \) and \( p(x|c_2)p(c_2) \), as a function of \( x \), on the same plot, clearly showing the optimal decision boundary (or boundaries).

3. Using integrals with specific limits, write down an equation for the Bayes error rate for this problem.

4. Evaluate the integrals in part 3 to estimate the Bayes error rate for this problem within 3 decimal places of accuracy (you will need to use numerical tables or use integration functions in Matlab or R to do this).

5. Answer parts 2 to 4 above but now with \( p(c_1) = 0.9 \).

Note that the Bayes error rate is defined as the minimum achievable (or optimal) classification error rate for a classification problem. For two classes, in the general case it is defined as

\[
\int_{\mathbf{x}} \min\{P(c_1|\mathbf{x}), 1 - P(c_1|\mathbf{x})\} \ p(\mathbf{x}) \ d\mathbf{x}
\]

Note that if the \( \mathbf{x} \) space is partitioned into some number of contiguous decision regions, such that within each region one of the classes is always more likely than the other, then the Bayes error rate can be re-expressed as a sum of integrals, one for each decision region. For 2-class problems, decision boundaries (between decision regions) are defined by the condition \( p(c_1|\mathbf{x}) = p(c_2|\mathbf{x}) = 0.5 \).

**Problem 6:**

Consider a classification problem with 2 classes and a single real-valued feature vector \( X \). For class 1, \( p(x|c_1) \) is uniform \( U(a,b) \) with \( a = 1 \) and \( b = 3 \). For class 2, \( p(x|c_2) \) is exponential with density \( \lambda \exp(-\lambda x) \) where \( \lambda = 1 \). Let \( p(c_1) = p(c_2) = 0.5 \).
1. Determine the location of the optimal decision regions

2. Draw a sketch of the two class densities multiplied by $P(c_1)$ and $P(c_2)$ respectively, as a function of $x$, clearly showing the optimal decision boundary (or boundaries)

3. Using integrals with specific limits, write down an equation for the Bayes error rate for this problem.

4. Evaluate the integrals in part 3 to compute the Bayes error rate for this problem within 3 decimal places of accuracy

5. Answer the questions above but now with $a = 0$ and $b = 2$.

Problem 7:

Consider a classification problem with 2 classes and where the $\mathbf{x} = (x_1, x_2)$ variable is 2-dimensional. For class $c_1$ the data is distributed uniformly between $(1, 4)$ and $(5, 8)$, and for class $c_2$ the data is distributed uniformly between $(4, 1)$ and $(8, 5)$. Let $p(c_1) = 0.6$ and $p(c_2) = 0.4$. Compute the Bayes error rate for each of the following situations, showing in a clear step-by-step manner (i.e., with appropriate integrals) how you derived your result:

1. Using only $x_1$ for classification

2. Using both $x_1$ and $x_2$ for classification

Problem 8:

Consider a linear classifier for a $K$-class problem defined by $K$ linear discriminant functions

$$g_k(\mathbf{x}) = w_k^T \mathbf{x} + w_{k0} \quad i = 1, \ldots, m.$$ 

Prove that the decision regions are convex by showing that if $\mathbf{x}_1 \in DR_i$ and $\mathbf{x}_2 \in DR_k$ then $\alpha \mathbf{x}_1 + (1 - \alpha) \mathbf{x}_2 \in DR_k$ if $0 \leq \alpha \leq 1$. The notation $\mathbf{x}_1 \in DR_k$ means that $\mathbf{x}_1$ lies within the decision region $DR_k$. 