CS 274A Homework 6

Probabilistic Learning: Theory and Algorithms, CS 274A, Winter 2014

Due Date: Wednesday March 12th in class

Problem 1: (10 points)

Consider a classification problem with 2 classes and a single real-valued feature vector X, where class 1 has a Gaussian density $p(x|c_1)$ with parameters μ_1 and σ_1^2, and class 2 has a Gaussian density $p(x|c_2)$ with parameters μ_2 and σ_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 $\sigma_1^2 = 1$ 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. Estimate the Bayes error rate for this problem within 3 decimal places of accuracy (you will need to use numerical tables for evaluating integrals of the Gaussian density function).

4. Answer parts 2 and 3 above but now with $p(c_1) = 0.9$.

You can use results/solutions from earlier homeworks if you wish.

Note that the Bayes error rate is defined as the minimum achievable (or optimal) classification error rate for a classification problem (assuming 0-1 loss). For two classes, in the general case it is defined as

$$\int_{\underline{x}} \min\{P(c_1|\underline{x}), 1 - P(c_1|\underline{x}\} \, p(\underline{x}) \, d\underline{x}$$

0.1   Solution to Problem 1

Use Multivariate Gaussian from EM Notes to get $p(x|c_1)$ and $p(x|c_2)$. Compute the following

$$p(c_1|x) = \frac{p(x|c_1)p(c_1)}{p(x|c_1)p(c_1) + p(x|c_2)p(c_2)}$$

Find when that is equal to 0.5 as that is the decision boundary. It occurs when

$$\frac{p(x|c_2)p(c_2)}{p(x|c_1)p(c_1)} = 1$$

This is the same as

$$\frac{p(x|c_2)}{p(x|c_1)} = \frac{p(c_1)}{p(c_2)}$$

Let

$$K_1 = \frac{1}{(2\pi)^{d/2}|\Sigma_1|^{1/2}}$$

$$K_2 = \frac{1}{(2\pi)^{d/2}|\Sigma_2|^{1/2}}$$

$$f_1(x) = (x - \mu_1)^T \Sigma_1^{-1} (x - \mu_1)$$

$$f_2(x) = (x - \mu_2)^T \Sigma_2^{-1} (x - \mu_2)$$

We now have to solve the following

$$\frac{K_2 \cdot e^{-f_2(x)/2}}{K_1 \cdot e^{-f_1(x)/2}} = \frac{p(c_1)}{p(c_2)}$$

Let

$$Q = \frac{p(c_1) K_1}{p(c_2) K_2} = \frac{p(c_1) |\Sigma_2|^{1/2}}{p(c_2) |\Sigma_1|^{1/2}}$$

Then we have to solve the following

$$e^{f_2(x) - f_1(x)} = Q^2$$

$$f_2(x) - f_1(x) = 2log(Q)$$

$$(x^T \Sigma_2^{-1} x - x^T \Sigma_1^{-1} x) - (\mu_2^T \Sigma_2^{-1} \mu_2 - \mu_1^T \Sigma_1^{-1} \mu_1) = 2log(Q)$$

Let

$$S = 2log(Q) + (\mu_2^T \Sigma_2^{-1} \mu_2 - \mu_1^T \Sigma_1^{-1} \mu_1)$$

$$M = \Sigma_2^{-1} - \Sigma_1^{-1}$$

Then we are finally left with solving

$$x^T M x = T$$

This is a quadratic form in d variables.

Problem 2: (10 points)

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 $a = 2$ and $b = 4$. 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. Compute the Bayes error rate for this problem within 3 decimal places of accuracy

4. Answer the questions above but now with $a = 2$ and $b = 22$.

0.2 Solution to Problem 2

From previous problem we have to solve

$$\frac{p(x|c_2)}{p(x|c_1)} \leq \frac{p(c_1)}{p(c_2)}$$

Given our values we have

$$\frac{p(x|c_2)}{p(x|c_1)} \leq 1$$

$$p(x|c_2) \leq p(x|c_1)$$

That is when it is optimal to choose c_1. This occurs when

$$exp(-x) \leq 0.5$$

$$x \geq log(2)$$

However, we can only choose c_1 for $2 \leq x \leq 4$. Our condition is satisfied in that region, thus the decision regions are as follows:

c_1 if $2 \leq x \leq 4$

c_2 otherwise

Problem 3: (10 points)

Consider a 2-class classification problem with d-dimensional real-valued inputs $\underline{x}$, where the class-conditional densities, $p(\underline{x}|c_1)$ and $p(\underline{x}|c_2)$ are multivariate Gaussian with different means $\underline{\mu}_1$ and $\underline{\mu}_2$ and a common co-variance matrix Σ, with class probabilities $P(c_1)$ and $P(c_2)$.

1. Write the discriminant functions for this problem in the form of $g_1(\underline{x}) = \log p(x|c_1) + \log p(c_1)$ (same for $g_2(\underline{x})$).

2. Prove that the optimal decision boundary, at $g(\underline{x}) = g_1(\underline{x}) - g_2(\underline{x}) = 0$, can be written in the form of a linear discriminant, $\underline{w}^t \underline{x} + w_0 = 0$, where $\underline{w}$ is a d-dimensional weight vector and w_0 is a scalar, and clearly indicate what $\underline{w}$ and w_0 are 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 $\underline{x}$ is assigned to the mean that it is closest to, and where distance is measured using Mahalanobis distance.

Problem 4: (10 points)

Consider the situation in Problem 3, for the case of $d = 2$ and $p(c_1) = p(c_2)$, and with the mean of class 1 at $(1, 1)$ and the mean of the second class 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. If you wish you can use MATLAB to generate the plots for you, but hand-drawn is also fine.

2. the location of the optimal decision boundary relative 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.

Problem 5: (10 points)

Consider a 2 class problem with labeled training data $D = \{\underline{x}_i, c_i\}$, where $c_i \in \{0, 1\}$ and $i = 1, \ldots, N$. Assume you are working with 2 different classification models M_1 and M_2 with parameters θ and ϕ respectively, where each model's functional form can be written as $p(c|\underline{x}, \hat{\theta})$ and $p(c|\underline{x}, \hat{\phi})$ respectively, where $\hat{\theta}$ and $\hat{\phi}$ are point estimates of the parameters. For example M_1 could be a logistic regression model and M_2 could be a neural network with a single hidden layer—the details of these models are not important for this problem. Assume you are fully Bayesian in the way you wish to approach the problem. How would you make classification decisions as a Bayesian, i.e., how would you decide whether the probability of one class or the other is more likely given a new test vector $\underline{x}$? (Hint: it may help to start by thinking about $p(c|\underline{x}, D)$). Clearly explain all steps in your answer.

Problem 6: (10 points)

In class we showed that least squares regression could be related to a conditional Gaussian likelihood. Assume we have training data in the form $D = \{x_i, y_i\}$, where x and y are both 1-dimensional and $i = 1, \ldots, N$. Say we assume a model of the form $E[y] = ax + b$ with y having Gaussian noise with variance σ^2 and mean $E[y]$. Derive maximum likelihood estimates for each of a, b, and σ^2.