CISC859 Pattern Recognition
Assignment 1

Handwritten answers are fine, but please write clearly enough so your work is easy to read.

Readings Be thorough in reading pages 1 to 27 of Duda, Hart, and Stork. Skim the rest of chapter 2, reading as much as you need to in order to answer the questions in assignments 1 and 2. The starred sections can be skipped, except for Section 2.7 on Error Probabilities and Integrals -- that is an important topic, so read this section carefully.

Most of Assignment 1 consists of introductory and background problems. If you find these fairly easy, that’s great! If not, then these problems provide you with needed introduction and review. Problem 8 is very important -- this problem asks you to apply the Bayes’ classifier in a simple case where the probability densities are assumed to be uniform. If you are confused by this problem, spend time on it, think carefully about it, discuss with other students in the class.

1) Consider a two-class problem, such as classifying a sample as “cat” versus “dog”. We define two features for this problem, and measure the feature-values obtained from various training samples. In the following plot, each c denotes the features measured for a cat sample, and each d denotes the features measured for a dog sample. Which feature has better discrimination power? Explain.

![Feature Plot]

2) We must classify a sample as “cat” or “dog”, but we cannot measure any features. Thus, we have to classify using some kind of guessing strategy. We know that P(cat) = 0.7, so 70% of the time the sample really is a cat.

2(a) What is P(dog)? [Reminder: The prior probabilities for all classes have to sum up to one.]

2(b) We try a strategy of guessing dog or cat with equal probability: 50% of the time we guess dog and the rest of the time we guess cat. (For example, we could flip a coin to decide which answer to give.) What is the probability of error, using this strategy?

2(c) We try a strategy of guessing dog and cat with the same frequency as these actually occur: 70% of the time we guess cat, and 30% of the time we guess dog. What is the probability of error, using this strategy?

2(d) We try a strategy of guessing cat 100% of the time. What is the probability of error, using this strategy?

2(e) Looking over your answers for 2(b) (c) and (d), what strategy is best? In other words, what strategy has the lowest probability of error?

3) Suppose P(cat) can have any value in the range 0 to 1. What is the probability of error, when we use the equal-probability guessing strategy from problem 2(b)? Justify your answer. [Hint: You should find that the probability of error is always 50%, no matter what the value of P(cat).]
Review of integration. To integrate a function that has discontinuities, break the integral into a sum of integrals, where each new integral covers a range in which the function is continuous. For example, in 4(a), you need to break the integral that ranges from $-\infty$ to 1 into a sum of two integrals “$-\infty$ to 0” and “0 to 1”. (One of these integrals evaluates to zero.) Similarly, in 5(a) and 5(b) you need to break an integral “$-\infty$ to $\infty$” into a sum of three integrals.

4) What is \[ \int_{-\infty}^{1} f(x) \, dx \]

(a) if \( f(x) = \frac{1}{2} \) for \( 0 \leq x \leq 2 \)
\( = 0 \) elsewhere

(b) if \( f(x) = \frac{8}{2} \) for \(-2 \leq x \leq -1 \)
\( = \frac{7}{8} \) for \( 9 \leq x \leq 10 \)
\( = 0 \) elsewhere

4(c) Are the functions \( f(x) \) in (a) and (b) probability density functions? To answer this, check whether they satisfy the definition of probability density function: \( f(x) \geq 0 \) for all \( x \), and \( \int_{-\infty}^{\infty} f(x) \, dx = 1 \).

Review of the definitions of mean and variance.

For a probability density function \( f(x) \): mean \( \mu = \int_{-\infty}^{\infty} x \, f(x) \, dx \) This is also called \( E[x] \), the expectation of \( x \).

\[
\text{variance } \sigma^2 = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) \, dx \quad \text{This is also called } E[(x - \mu)^2]
\]

5) For 5(a) to 5(c), keep in mind that when someone says “density \( p(x) \) is uniform on interval \([a, b]\)”, this means that \( p(x) = \frac{1}{b-a} \) for the range \( a \leq x \leq b \) and \( p(x)=0 \) elsewhere.

5(a) Density \( g(x) \) is uniform on the interval \([10, 15]\). What is the mean of \( x \), when \( x \) is distributed according to \( g(x) \)? For practice, compute this two ways. First by “inspection”: for a uniform density, the mean is the middle value in the range 10..15. Second, compute the mean by evaluating the integral given in the above definition. Since \( g(x) \) is a discontinuous function, you must restate the integral as a sum of several integrals, each of which covers a range where the function is continuous. (Three integrals are needed here, for the ranges \(-\infty\) to 10, 10 to 15, and 15 to \( \infty \). Two of these integrals evaluate to zero.)

5(b) What is the variance of \( x \), when \( x \) is distributed according to \( g(x) \)? The variance is hard to figure out by “inspection”. Use the integral given in the definition above. Again, you need to restate this as a sum of three integrals (two of which evaluate to zero). To simplify the calculation, try replacing \( g(x) \) by a shifted density \( h(x) \), where \( h(x) \) is uniform on the interval \([-2.5, 2.5]\). (The variance of \( g(x) \) and \( h(x) \) is the same; make sure you understand why.)

5(c) Density \( d(x) \) is a uniform density, with mean \( \mu=2 \) and variance \( \sigma^2 = 1/3 \). You know that \( d(x) \) is uniform over some range \([a, b]\), but you aren’t explicitly told the values of \( a \) and \( b \). Use the information about mean and variance to calculate the values \( a \) and \( b \).

Hint: One way to figure this out is by considering a zero-mean uniform distribution (uniform on interval \([-c, c]\)), and computing what the variance \( \sigma^2 \) would be as a function of \( c \). Then solve for \( c \) in terms of \( \sigma \).
6) This problem reviews the Normal Density, also called the Gaussian density, which is used to model many physical processes. The Normal density is a “bell-shaped curve” characterized by two parameters: the mean $\mu$ and the variance $\sigma^2$. As an example: suppose I want to fit a Normal density to the set of marks in my undergraduate class. I need to find the $\mu$ and $\sigma$ values for the bell-shaped curve which best matches my data:

Finding $\mu$ and $\sigma$ is called “parameter estimation”, a topic covered in Chapter 3 of DHS. Of course, I can only get a good fit to my data if it actually has a bell shape. If half the class failed and half the class got an A then my data has a bimodal distribution, and no choice of $\mu$ and $\sigma$ is going to give a very good fit.

Informally, think of the Normal density as a bell-shaped curve with mean $\mu$ and with 95% of the samples falling within 2$\sigma$ of the mean. The equation for the Normal density, in terms of the mean $\mu$ and variance $\sigma^2$, is given on page 32 of DHS:

$$p(x) = \frac{1}{\sqrt{2\pi} \sigma} e^{-\frac{1}{2} \left( \frac{x-\mu}{\sigma} \right)^2}$$

Warning: due to that nasty $e^{-x^2}$ term there is no closed-form solution to $\int_{-\infty}^{K} p(x) \, dx$. The answer for particular values of $K$ can be looked up in published tables (often titled “the error function”), which list the numeric value of the integral for selected $K$ values. These table entries are computed by numerical integration, where a computer program adds up the areas of many tiny rectangles, which together fill the area under the curve.

Assignments in this course mainly use two types of densities: uniform densities because these are mathematically so easy (they make good exercises), and Normal densities because these are useful for modeling random processes in the real world.

6(a) Sketch the normal density $p(x)$ for $\mu=0$ and $\sigma=1$. (To do this, find the value of $p(x)$ for $x=\mu$; this is the value at the mean, and gives you the “peak” of the bell curve. Then pick one or two other value of $x$ to get additional points on your sketch (symmetric about the mean), and then draw a curve.)

6(b) Sketch the normal density $p(x)$ for $\mu=3$ and $\sigma=2$.

7) Suppose $x$ is a random variable drawn from a density $p(x)$ that is uniform in the range $[0,4]$. This means that $p(x)$ is $1/4$ for $0 \leq x \leq 4$ and $p(x)$ is zero elsewhere. We get 4 independent observations of $x$. What is the probability that all four observations are in the range $[0, 3]$? (If you are stuck: start off by figuring out what is the probability that the first observation is in the range $[0, 3]$.)
8) Consider a two-class, single-feature problem where the underlying class densities are uniform, and the prior probabilities are equal, i.e., \( P(\omega_1) = P(\omega_2) = 1/2 \). Since \( p(x | \omega_i) \) is uniform for classes 1 and 2, we can write (for \( i=1 \) and \( i=2 \)):

\[
p(x | \omega_i) = \frac{1}{(H_i - L_i)} \text{ for } x \text{ in the interval } [L_i, H_i] \\
= 0 \text{ elsewhere}
\]

For each case (a) to (d), do three things:

(i) sketch the two functions \( p(x | \omega_i) \)
(ii) define an optimal classification strategy phrased as “Guess \( \omega_1 \) if \( x \) is in the range <whatever>”
(iii) compute \( P(\text{error}) \), the probability of classification error for classification strategy (ii)

8(a) \( L_1 < H_1 < L_2 < H_2 \)
8(b) \( L_1 < L_2 < H_1 < H_2 \) and \( H_1 - L_1 = H_2 - L_2 \)
8(c) \( L_1 < L_2 < H_1 < H_2 \) and \( H_1 - L_1 > H_2 - L_2 \)
8(d) \( L_1 < L_2 < H_2 < H_1 \) and (by definition) \( H_1 - L_1 > H_2 - L_2 \)

Advice: Make sure that you can solve this problem both through intuition and by applying the formulas for Bayes’ classifier and \( P(\text{error}) \). Because uniform densities are so easy to reason about and so easy to integrate, you can use this problem to make sure you understand the Bayes’ classifier at an intuitive level. That gives you a solid foundation for solving problems in which Bayes’ classifier is applied to Normal densities. (Note that uniform densities are rarely used in practical applications, because they generally do not provide a good fit for modeling phenomena that arise in nature or in computer science research. In contrast, normal densities are suitable for modeling a variety of natural phenomena.)

Students often make mistakes in computing \( P(\text{error}) \) in terms of \( L_1 H_1 L_2 H_2 \). Check your answer for reasonableness by trying out various values of \( L_1 H_1 L_2 H_2 \). What are the minimum and maximum values of \( P(\text{error}) \) you can get? Since probabilities have to be between 0 and 1, you certainly made a mistake if your \( P(\text{error}) \) expression can produce a negative value or a value larger than 1.0. Furthermore, if your \( P(\text{error}) \) expression can produce a value between 0.5 and 1.0, you can also conclude that your expression is incorrect. This is because random guessing gets the right answer half the time in a two-class problem, as we saw in problem 3 of this assignment. Since the Bayes’ classifier is provably optimal – meaning that it has the lowest possible \( P(\text{error}) \) rate among all possible classifiers – it must be the case that \( P(\text{error}) \leq 0.5 \).