CISC859 Pattern Recognition, Winter 2019

Assignment 1, due in class on January 16

Assignments can be completed individually, or in a group of two or three students – your choice. Answers can be handwritten or typed, whatever you find easiest.

This first assignment provides introduction and review. If you find the questions easy, that's great! If not, then review as needed, discuss with other students in the class, see me in office hours if you need extra help.

Readings

Introduction to pattern recognition

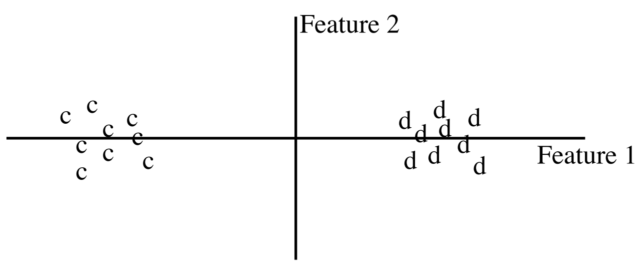
- DHS (textbook by Duda, Hart, and Stork) Chapter 1
- Course reader pages 1-4

Introduction to Bayes classifier

- DHS Section 2.1
- Course reader pages 8-16

Feature space

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



Classification using only prior probabilities

2) Classify a sample as *cat* or *dog* without measuring any features, given the information that P(cat) = 0.7 so 70% of the time the sample is a cat.

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

2(b) Consider the 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 when this strategy is used?

2(c) Consider the 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 when this strategy is used?

2(d) Consider the strategy of guessing cat 100% of the time. What is the probability of error?

2(e) Which of the strategies 2(b) (c) and (d) is best, having the lowest probability of error?

3) Generalize problem 2 where P(cat) was known to be 0.7. Now P(cat) can be any value in the range 0 to 1. What is the probability of error for the equal-probability guessing strategy defined in problem 2(b)? Justify your answer.
Hint: You should find that P(error)=50%, no matter what the value of P(cat).

Review: Integration of discontinuous functions

To integrate a function that has discontinuities, break the integral into a sum of integrals where each one covers a range in which the function is continuous. For example, break the integral in 4(a) into a sum of two integrals $-\infty$ to 0 and 0 to 1. (One of these integrals evaluates to zero.) You can use this problem as a review of integration; it is also easy to figure out the answer without integration because here the area under the curve has a rectangular shape.

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

(a) if $f(x) = 1/3$ for $0 \le x \le 2$
$= 0$ elsewhere

(b) if $f(x) = 1/8$ for $-2 \le x \le -1$
$= 7/8$ for $9 \le x \le 10$
$= 0$ elsewhere

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

Review: Mean and variance

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

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

Standard deviation is the square root of variance. Notice that *standard deviation* is analogous to *Euclidean distance*: both are calculated as a square root of a sum of squared distances. In a 1D Normal density, 95% of the samples are within 2σ of μ.

Reminder for problem 5: "density p(x) is uniform on interval $[a, b]$" means that $p(x) = 1/(b-a)$ for $a \le x \le b$ and $p(x) = 0$ elsewhere.

5(a) Density g(x) is uniform on the interval [10, 15]. What is the mean of g(x)? For practice, figure out the mean in two ways and make sure you get the same answer:

- By inspection. For a uniform density the mean is the middle value in the range, here 10..15.
- By evaluating the integral in the above definition. Since g(x) has two discontinuities, you must break the integral into a sum of three integrals, for the ranges $-\infty$ to 10, 10 to 15, and 15 to ∞. Two of these integrals evaluate to zero.

5(b) What is the variance of g(x)? The variance is hard to figure out by inspection, so evaluate the integral given in the above definition. Again the two discontinuities mean that you have to break the integral into a sum of three integrals, two of which evaluate to zero.
Hint: To simplify the evaluation of the integral, replace g(x) by a shifted density h(x) that is uniform on the interval [-2.5, 2.5]. The densities g(x) and h(x) have the same variance; make sure you understand why.

5(c) Density d(x) is a uniform density with mean $\mu=2$ and variance $\sigma^2=1/3$ (so the standard deviation is about .58). This density d(x) is uniform over some range $[a, b]$. Your job is to use the given mean and variance to calculate the values a and b.
Hint: Make life easier for yourself by first considering a zero-mean uniform density: a density that is uniform on the interval $[-c, c]$. Plug this density into the definition of σ^2 to derive a formula for the value of σ^2 as a function of c. This allows you to calculate c by using the given variance $\sigma^2=1/3$. Next you can find values a and b by shifting the density to make the mean equal to 2.

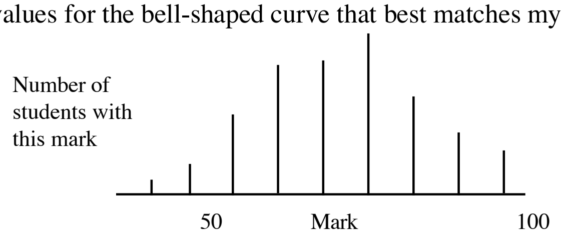
Review of the Normal Density

Here is the equation for the Normal density, also called the Gaussian density, from DHS page 32. There are two parameters: the mean μ and standard deviation σ (variance σ^2).

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

The Normal density is a bell-shaped curve with mean μ and with 95% of the samples falling within 2σ of the mean. It is impossible to symbolically integrate this density: due to the e^{-x^2} term there is no closed-form solution for $\int_{-\infty}^{K} p(\mathrm{x})\,d\mathrm{x}$ Instead, the value of the integral (for various values of K) can be looked up in published tables; look for tables of "the error function". The values in the tables are computed by numerical integration, where software computes the sum of the areas of many tiny rectangles that together fill the area under the curve (in the limit, as the areas of the tiny rectangles go to zero and the number of tiny rectangles goes to infinity).

Assignments in this course mainly use two densities: (1) uniform densities because these are mathematically simple and therefore make good introductory exercises, and (2) Normal densities because these are useful for modeling many random processes in the real world. As an example: suppose I want to fit a Normal density to a set of course marks. I need to find μ and σ values for the bell-shaped curve that best matches my data:



Finding μ and σ is called *parameter estimation*, a topic covered in DHS Chapter 3. A Normal density can be a good fit to my data only if my data has approximately a bell shape. If half the class failed and half the class got an A then the data has a bimodal distribution, and no choice of μ and σ is going to result in a Normal density that is a good fit to this data.

6(a) Sketch the normal density $p(\mathrm{x})$ for μ=0 and σ=1.
To do this, find the value of $p(\mathrm{x})$ for x = μ; this is the value at the mean, and gives you the "peak" of the bell curve. Pick one or two other values of x to get additional points on your sketch (symmetric about the mean), and then draw a curve.

6(b) Sketch the normal density $p(\mathrm{x})$ for μ=3 and σ=2.

Random variables

7) Suppose x is a random variable drawn from a density p(x) that is uniform in the range [0..4]. We get 4 independent observations of x. What is the probability that all four observations are in the range [0, 3]?
Hint: as a first step, figure out the probability that the first observation is in the range [0, 3].

Introductory problem for the Bayes Classifier

Here you apply the Bayes classifier in a simple situation: only two classes, one feature, and uniform probability densities. Assignment 2 gives you more practice with the Bayes classifier.

8) Consider a two-class, single-feature classification problem where the class-conditional probability densities are uniform. The density for class 1 is uniform in the range 0 to 10 and the density for class 2 is uniform in the range 8 to 13.

(a) The value of $p(x \mid \omega_1)$ is zero outside the range $0 \le x \le 10$. What is the value of $p(x \mid \omega_1)$ in the range $0 \le x \le 10$?

(b) The value of $p(x \mid \omega_2)$ is zero outside the range $8 \le x \le 13$. What is the value of $p(x \mid \omega_2)$ in the range $8 \le x \le 13$?

(c) Sketch the two functions $p(x \mid \omega_i)$. Your sketch is analogous to the one shown in DHS Figure 2.1, but with uniform densities.

(d) Assume the prior probabilities are equal, i.e. $P(\omega_1) = P(\omega_2) = 1/2$.

(i) Define an optimal classification strategy. Something along these lines: "Classify the sample as ω_1 if x is in the range <whatever>, otherwise as ω_2."

(ii) Compute P(error), the probability of making an error when using this classification strategy. Hint: Because these are uniform densities, it is easy to figure out P(error) by inspection. What percentage of ω_1 samples are misclassified, and what percentage of ω_2 samples are misclassified? Combine those two cases to come up with the overall P(error), keeping in mind that half the samples are ω_1 samples and the other half are ω_2 samples.

(e) Repeat part (d), but this time use the prior probabilities $P(\omega_1)=0.8$ and $P(\omega_2)=0.2$