CISC859 Pattern Recognition, Winter 2019

Assignment 2, due January 28

Readings

- DHS Chapter 2
  Section 2.1 was in the readings for assignment 1. Read the rest of DHS chapter 2, putting your focus on the topics needed to answer the assignment questions. Study Section 2.7 on Error Probabilities and Integrals, but skip all the other sections that are marked with a * (that's 2.3.1, 2.3.2, 2.8. 2.10, 2.11, 2.12).
- Course reader pages 17-34. Minimum risk classifier, normal density, estimating P(error)

Bayes classifier

1) Recall that in problem 8 of assignment 1 you applied Bayes classifier to specific densities like "uniform in the range 0 to 10". Here generalize to densities "uniform in the range L_1 to H_1".

A two-class, single-feature problem has equal prior probabilities $P(\omega_1) = P(\omega_2) = 1/2$. The densities $p(x \mid \omega_i)$ are uniform, so we know that for i=1 and i=2: $p(x \mid \omega_i) = 1/(H_i - L_i)$ for x in the interval $[L_i, H_i]$
$= 0$ elsewhere

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

(i) sketch the two functions $p(x \mid \omega_i)$
(ii) define an optimal classification strategy "Classify the sample as ω_1 if x is in the range <whatever>"
(iii) compute P(error), the probability of classification error when using classification strategy (ii)

a) $L_1 < H_1 < L_2 < H_2$

b) $L_1 < L_2 < H_1 < H_2$ and $H_1 - L_1 = H_2 - L_2$

c) $L_1 < L_2 < H_1 < H_2$ and $H_1 - L_1 > H_2 - L_2$ Check your work by setting $L_1=0$ $H_1=10$ $L_2=8$ $H_2=13$. This should give you the answer for assignment 1 problem 8d.

d) $L_1 < L_2 < H_2 < H_1$ and (by definition) $H_1 - L_1 > H_2 - L_2$

Check your P(error) expressions. Using various values of L_1 H_1 L_2 H_2, what are the minimum and maximum possible values of P(error) according to your expression? Clearly you made a mistake if your P(error) can be less than zero or larger than one. Actually, any P(error) > 0.5 indicates a mistake: the Bayes' classifier is optimal and therefore outperforms the P(error)=0.5 attained by random guessing (recall problem 3 in assignment 1).

2) Consider a two-class, single-feature problem with the following class densities. Note that the first density is normal (bell-shaped curve) and the second one is uniform (rectangular shape).

$p(x \mid \omega_1)$ is **Normal** with mean $\mu_1=0$ and $\sigma_1^2 = 1$
$p(x \mid \omega_2)$ is **uniform** with mean $\mu_2=2$ and $\sigma_2^2 = 1/3$
As you showed in assignment 1 problem 5(c): $p(x \mid \omega_2)$ is uniformly distributed in the range [1, 3]

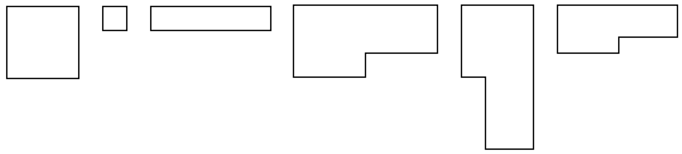
a) Sketch $p(x \mid \omega_1)$ and $p(x \mid \omega_2)$. Sketch both functions on the same plot (as is done in Figure 2.1 of DHS), so that you can compare them. A rough sketch is fine. My intention is that you do this by hand, but you can use a plotting package if you prefer.

b) Assume $P(\omega_1) = P(\omega_2) = 1/2$. State the classification strategy that minimizes classification error; describe the classification strategy in terms such as "classify as ω_1 when x is in the range <something>". Also write an expression for P(error) for this decision rule. Since there is no closed-form solution for integration of the Normal density, leave your answer in the form of an integral. If you wish, you can look up the answer to the integral in a published table.

c) Assume $P(\omega_1) = 3P(\omega_2)$. State the classification strategy that minimizes classification error. Write an expression for P(error) for this decision rule.

Features for classifying shapes

3) Consider the problem of recognizing 2D shapes in image data. For simplicity, consider polygonal shapes where all edges are parallel to one of the coordinate axes, as in these examples.



In many applications, rotation and scaling should not affect shape classification. For example, consider a system that recognizes parts as they come down an assembly line. The system should be able to recognize the parts at any orientation. (In a factory setting, scale invariance might not be needed, because we can control the distance from the camera to the assembly line.) In the above examples, shape 2 is a scaled version of shape 1, shape 6 is a scaled version of 4, shape 5 is a rotated version of 4.

(a) The following feature is proposed: $\text{Feature}_1 = \dfrac{\text{region\_area}}{\text{region\_perimeter}^2}$

Is Feature_1 *scale invariant*? In other words, if a polygon has $\text{Feature}_1 = F$, then does a scaled version of that polygon also have $\text{Feature}_1 = F$? Briefly justify your answer.

Is Feature_1 *rotation invariant*?

Is Feature_1 sufficient for discriminating among all possible different shapes? If so, argue why. If not, give an example of two different shapes that have the same value for Feature_1.

(b) Propose one or two other features that are scale and/or rotation invariant. Discuss their discrimination power.

(c) *Doing part (c) is **optional**; I encourage you to have a quick look at it.*
Is it possible to define one really powerful feature that can distinguish between all polygons? To state the problem more precisely, assume that a polygon is defined by a sequence of vertex locations (x_i, y_i), where x_i and y_i are real numbers. The question is whether it is possible to define one real-valued feature that uniquely encodes a polygon. Here we are not concerned with shifting or scaling or rotation: two polygons are supposed to map to different feature values if they don't have identical vertex locations. Can you find a way to define a real-valued feature F such that

if polygon A is different from polygon B, then $F(A) \neq F(B)$

Alternatively, can you prove (or just informally argue) that such a feature F cannot be defined?
Note that this result does not apply to computers. Computers cannot store real numbers: that would require an infinite amount of memory because real numbers have infinite precision. Our computers use floating-point numbers with finite precision.
Note for students who like fractals: you can also try to answer this question for fractal polygons -- these have an infinite number of vertices.

Nearest neighbor and Bayes classifiers in DHS toolbox

4) Look at the document Introduction to the DHS toolbox available on the course website under Pattern Recognition Resources http://research.cs.queensu.ca/~blostein/859_DHS_Toolbox.pdf. The third screenshot in this document shows the result of applying the Bayes' classifier and the 3-Nearest-Neighbour classifier to the clouds data set.

(a) As shown in the screen shot, the error rate for the Bayes classifier is 0.1 (i.e. 10%), whereas the 3-Nearest-Neighbour classifier has error rate 13% on the test set and 8.3% on the training set. So on the training set the 3NN classifier is outperforming Bayes. Can this be right??? Recall that the Bayes classifier is optimal when $p(\mathbf{x} \mid \omega_i)$ and $P(\omega_i)$ are known correctly, and that is the case here. The clouds data is generated from $p(\mathbf{x} \mid \omega_i)$ that are defined to be the sum of one or more Gaussians, and the Bayes' classifier is given the correct formulas for $p(\mathbf{x} \mid \omega_i)$.

(b) I mentioned in class that for a 1-Nearest-Neighbor classifier (often abbreviated to "nearest neighbor classifier"), the error rate on the training set is zero. Describe why this happens.

Analyzing the minimum-error-rate Bayes classsifier

Problems 5 and 6 are from the DHS textbook. Working on these problems is a great way to deepen your understanding of the Bayes classifier. I know that these problems might take you a fair bit of time, and I will not put problems of this type on the exam.

5) DHS page 68, problem 12.
Unfortunately, different printings of the textbook have different page numbers. So look for the problem statement that starts like this:

> Let $\omega_{max}(\mathbf{x})$ be the state of nature for which $P(\omega_{max}|\mathbf{x}) \geq P(\omega_i|\mathbf{x})$ for all i, i=1,...,c.

Hint: Keep in mind that the value of the subscript "max" depends on **x**. Get an intuitive understanding of the result that $P(\text{error}) \leq (c-1)/c$. The error rate for a classifier that uses random guessing is $(c-1)/c$. Under what conditions would a Bayes' classifier get such a high error rate?

Reject option in minimum-risk Bayes classifier

6) DHS page 68, problem 13.
This problem statement starts like this:

> In many pattern classification problems one has the option either to assign the pattern to one of *c* classes, or to *reject* it as being unrecognizable.

Hints: Find the risk for each action and pick the minimum risk. It's best to calculate the risk for actions 1..c (classify as class 1 to c) separately from the risk for action c+1 (reject). You can use the expression for risk given at the end of page 17 in the course reader. The risk for action c+1 should end up equal to λ_r and the risk for actions 1 to c should be an expression that involves λ_s. Compare these two expressions to see which one is smaller; this lets you derive the expression given in the problem statement: "$1 - \lambda_r / \lambda_s$".