

UNIVERSITY OF ILLINOIS AT URBANA-CHAMPAIGN
Department of Electrical and Computer Engineering

ECE 544 PATTERN RECOGNITION

Problem Set 1
Fall 2007

Issued: Tuesday, August 27, 2007

Due: Thursday, September 6, 2007

Reading for problem set 1: DHS (Duda, Hart, & Stork) chapter 2

Written Problems (40%)

DHS problems 2.6, 2.22, 2.30, 2.43.

Library Problem (20%)

Choose an application of pattern recognition that is interesting to you. Read at least one long peer-reviewed paper describing an approach that other researchers have taken to solve your problem or a related problem (“long” means 7-8 pages or more – e.g., journal articles, or long conference papers at conferences such as NIPS, or two short conference papers). Write a one-page summary of the article: (a) what was the application? (b) what methods have previously been used to address that application? (c) what specific problem with previous methods are the authors trying to solve? (d) what method did they propose? (e) how did they evaluate their proposal? (f) is the improvement statistically significant (i.e., not chance)? is the improvement worth the added complexity or cost?

At the end of your one-page summary, provide a bibliography, including at least one citation (the paper you are reviewing), in IEEE citation style.

Prescribed Computer Problem (20%)

DHS computer exercise 2.2. Your homework response should contain the following sections:

- (a) Narrative description of the experiment. Write a one-page description of your experimental methods and results.
- (b) Figures. Include all plots and figures requested by the problem statement. These figures should be described by figure number in your narrative (e.g., “Fig. 1 shows...”). Figures will only be read if they are described, by figure number, in the narrative.
- (c) Appendix A: Provide up to one page of pseudo-code or flowchart, describing how you implemented part or all of the required problem. This page should either be self-documenting (via comments), or fully described in the narrative, or both.
- (d) Appendix B: Provide up to one page of actual matlab (or other) code, showing how you implemented one or more of the ideas presented in Appendix A.

Grading criterion: this problem will be graded based on my assessment, after reading your Narrative, Appendix A, and Appendix B, of your ability to solve all of the technical problems posed in the problem statement. If I am unable to understand your methods or your results, your grade may suffer—in other words, clarity of thought and expression matter.

Collaborative problem-solving: you are allowed to work with other students. If you work with other students, you must (a) specify, in your narrative, the names of the other students that you worked with, and (b) write your own narrative summary, and your own Appendix A. I expect that all of the students in an informal study group may submit identical versions of Appendix B (or they may not—you need not choose, for your appendix, the same subroutine that your group-mates chose), but that you will not submit identical copies of the narrative or of Appendix A. In such cases, grades will be assigned primarily based on the clarity of your narrative and of your Appendix A.

Open-Ended Computer Problem (20%)

Find a source for labeled data in an application area that interests you. Choose a two-class subset of your data, with feature vectors that are at least three-dimensional. Repeat computer exercise 2.2 using these data. Report on your results using the same format that you used above (including narrative, figures, appendix A, and appendix B). You are allowed to work with others, under the same restrictions that apply to the prescribed computer problem.

Design your dichotomizer under whichever of the following assumptions is least inappropriate: (a) the data are Gaussian distributed, or (b) the data are discrete, or (c) if you know of a PDF that represents these data better than the Gaussian or the discrete PDFs (something not yet covered in class), you may use it.

You may find that performance on real-world data is not as good as performance on the canned data in the textbook. In order to convince me that your code works, you may provide any of the following figures. Note that you aren't *required* to turn in any figures at all—only those that will convince me of your ability to solve all relevant problems.

Optional but possibly useful plots: (a) histograms $p(x_1|\omega_1)$ and $p(x_1|\omega_2)$ of the first feature given the two classes, (b) the Gaussian PDFs or discrete PMFs that you used to approximate $p(x_1|\omega_1)$ and $p(x_1|\omega_2)$ for the purpose of designing a classifier, and (c) the discriminant function, perhaps with the decision threshold drawn.