SD 372 Intro. to Pattern Recognition April 9, 2001

Final Examination

Professor Paul Fieguth

Aids Permitted: *Two* 8.5x11 pages, *No* calculator.
Please turn off cellphones, pagers, and other pointless electronic gadgets etc.

*** Well-drawn sketches / diagrams can be very helpful. ***

The grade value for each question is indicated in brackets [ ] next to the question number. I *will* give part marks for relevant statements or insights.

Manage your time!!! This is the most common problem which I see among students.

[2%] Bonus Question

Do **NOT** waste your time here unless you are happy with your answers to the rest of the exam!!!

I have no idea where I came across the following. I had to evaluate this as part of some grad school assignment. I did it using a calculator, but it turns out you can solve it analytically. Reduce the following to a simple, analytic expression:

$$\left(\frac{3}{(1+\sqrt[3]{2})^2}+1\right)^{\frac{1}{2}}$$

[22%] **1. MICD**

GIVEN: Two clusters in 2D space $\underline{x}$ with means and covariances as follows:

$$\underline{\mu}_A = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \quad S_A = \begin{bmatrix} 2 & 0 \\ 0 & 1 \end{bmatrix} \qquad \underline{\mu}_B = \begin{bmatrix} 3 \\ 2 \end{bmatrix} \quad S_B = \begin{bmatrix} 4 & -2 \\ -2 & 4 \end{bmatrix}$$

a) Sketch the unit standard-deviation ellipse for class A.

b) Sketch the unit standard-deviation ellipse for class B.

c) Draw a sketch showing the two classes and the MED classification boundary.

d) Similarly sketch (but don't try to compute) the MICD classification boundary.

e) Use Fisher's method to find the best feature direction $\underline{w}$.

f) Let our new feature be $y = \underline{w}^T \underline{x}$. Find the mean and variance of classes A and B on y.

g) Sketch (don't try to compute) the MICD classification boundary on y.

h) Note that a classifier on y implies a classifier on $\underline{x}$. That is, for each $\underline{x}$ we find $y = \underline{w}^T \underline{x}$ and then we determine the class based on y. So sketch the classification boundary of your MICD classifier from (g), but sketch it back in the 2D domain of $\underline{x}$, together with the unit standard-deviation ellipses for A and B.

[25%] **2. Parameter Estimation**

Every student knows that exam grades are pretty variable – you have good days and bad days etc. Let's suppose that a students "intrinsic" ability is represented by an unknown α, where α is measured in percent. But the grade g (also measured in percent) that a student gets on an exam (a "measurement") is Gaussian distributed around the ability α:

$$p(g|\alpha) = \frac{1}{\sqrt{2\pi}10} \exp\left\{-\frac{(g-\alpha)^2}{2 \cdot 10^2}\right\}$$

a) What is the probability that a student of ability α will get a grade below 50% on an exam?

b) Normally we try to get a better idea of a student's ability by having them write several tests. Suppose a student writes N tests, getting grades $g_1, \ldots, g_N$, and the grades are independent.

   What is $\hat{\alpha}_{ML}(g_1, \ldots, g_N)$, the Maximum Likelihood estimate of the student's ability?

c) What is the probability, as a function of α, that $\hat{\alpha}_{ML}(g_1, \ldots, g_N)$ is below 50%?

d) It is interesting to assess how more and more tests (larger N) affect the precision of the estimated $\hat{\alpha}$. Find the limit (showing your work) of

$$\lim_{N \to \infty} \Pr(\hat{\alpha}_{ML}(g_1, \ldots, g_N) < 50)$$

   first for the case $\alpha = 49$, and secondly for the case $\alpha = 51$.

Maybe a Gaussian distribution isn't such a good model for how an exam grade g is distributed, because it seems more likely that one would have a bad day, or accidentally mess up a question, than to accidentally get a whole lot of questions right. Maybe it is more reasonable to say that the actual grade g can never be higher than α. For the rest of this question, suppose we model g as

$$p(g|\alpha) = \begin{cases} 0 & g > \alpha \\ \frac{1}{10} \exp\left\{\frac{g-\alpha}{10}\right\} & g \leq \alpha \end{cases}$$

e) Draw a sketch of $p(g|\alpha)$ for $\alpha = 75$.

f) What is the probability that a student of ability α will get a below 50% on an exam?

g) As in part (b), suppose a student writes N tests, getting grades $g_1, \ldots, g_N$, and the grades are independent.

   What is $\hat{\alpha}_{ML}(g_1, \ldots, g_N)$, the Maximum Likelihood estimate of the student's ability?

[28%] **3. NonParametric Estimation:**

Suppose we have a one-dimensional non-parametric estimation problem. We're given N samples $x_1, \ldots, x_N$.

a) Write down the general form for $\hat{p}_{hist}(x)$ for the histogram method.

   Show that $\int \hat{p}_{hist}(x)dx = 1$; that is, show that the estimated PDF is normalized. Note: you don't need to make any assumptions about number of bins or bin widths.

b) Write down the general form for $\hat{p}_{parz}(x)$ for the Parzen method.

   Show that $\int \hat{p}_{parz}(x)dx = 1$; that is, show that the estimated PDF is normalized. Note: you don't need to make any assumptions about the type of window shape or its width.

c) Briefly describe how one computes $\hat{p}_{knn}(x)$ for the kNN method.

   Show that $\int \hat{p}_{knn}(x)dx > 1$; that is, show that the estimated PDF is never normalized, no matter what the data points $x_1, \ldots, x_N$ are.

d) Briefly describe the advantages of the Parzen method over histograms or kNN. That is, why do we normally prefer Parzen?

Now suppose we have two equally likely classes and a total of four data points in 1D:

$$\begin{array}{ll} \text{Class A:} & x_1 = -3,\ x_2 = -1 \\ \text{Class B:} & x_3 = +1,\ x_4 = +2 \end{array}$$

We want to estimate $p(x|A), p(x|B)$ from the four data points using the Parzen method, with a Gaussian window having a standard deviation of 1.0

e) Draw a neat sketch of $\hat{p}(x|A), \hat{p}(x|B)$.

f) Now suppose we use $\hat{p}(x|A), \hat{p}(x|B)$ to develop a ML classifier. How many classification boundaries will there be? Why?

   Clearly there will be a boundary between the two clusters, near 0. Will the boundary be at zero, slightly below zero, or slightly above? Why?

g) Suppose I propose the following classifier:

$$\text{Say} \quad \begin{array}{ll} x \in A & \text{if } x < 0 \\ x \in B & \text{if } x \geq 0 \end{array}$$

   Based on $\hat{p}(x|A)$ and $\hat{p}(x|B)$, what is the probability of classification error $P(\epsilon)$?

[25%] **4. Prototypes**

We want to explore different choices of prototypes.

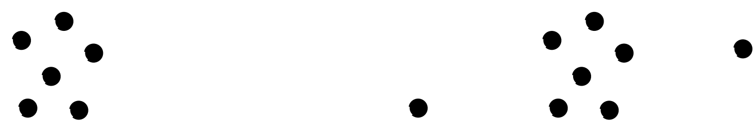
a) Define a "prototype." For what sorts of Pattern Recognition problems do we care about prototypes?

b) True or False: As k increases, the kNN prototype converges to the mean prototype.

c) True or False: As k increases, the kFN prototype converges to the mean prototype.

d) For the kNN or kFN prototypes, what are the tradeoffs in varying k? That is, what are the pros and cons associated with choosing a large or a small value of k?

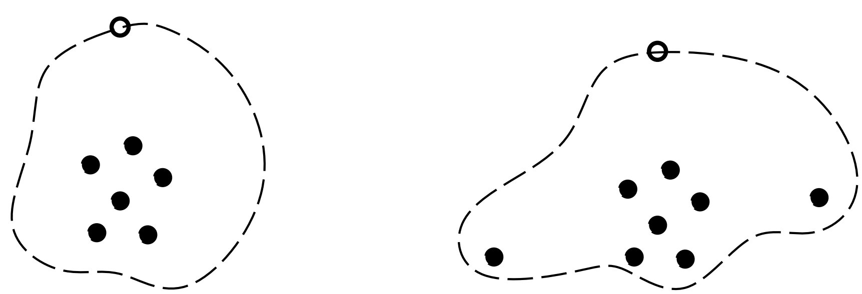
e) Consider the two clusters below:



Each solid dot is one data point. Note that the clusters are the same, except that the right one has two additional noisy points.

Let our definition of distance be the normal Euclidean (standard "blackboard") distance.

Then for every sort of prototype definition, we can draw a constant-distance contour for the cluster, for example

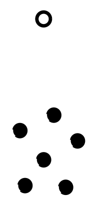


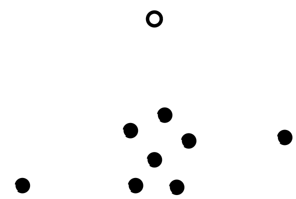
Obviously we'll have a contour for every possible distance. Draw constant distance contours passing through the open circle (as I did in the above examples; the open circle is not a data point) for both of the clusters for each of the following five prototype definitions:

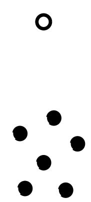
1. Mean
2. FN
3. NN
4. 2NN
5. 4NN

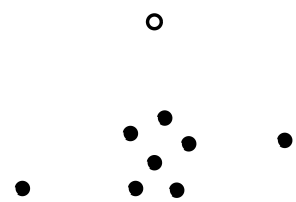
The following two pages provide templates that you may find easier to use than redrawing the clusters in your exam book.

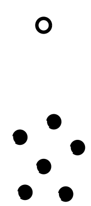
You may find the following helpful for answering Question 4.
Make sure you clearly label what each plot is answering.
Make sure your constant-distance contours pass through the open circle.

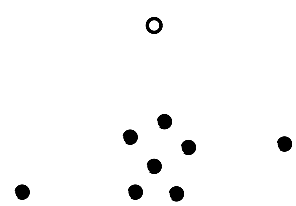


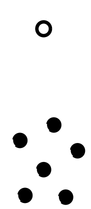


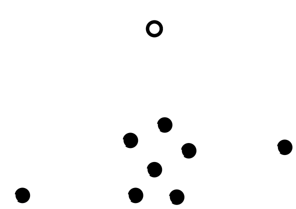


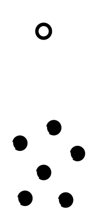


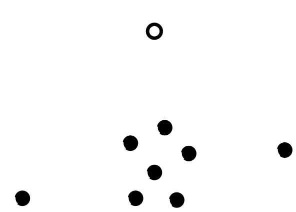












A second page (you may or may not need it).
Make sure you clearly label what each plot is answering.
Make sure your constant-distance contours pass through the open circle.