CARLETON UNIVERSIY
SCHOOL OF COMPUTER SCIENCE
COMP 5107 (Winter 2020)
PATTERN RECOGNITION
NEWS

January 7, 2020
1. Classes start.
2. Hope you all enjoy the course!

March 14, 2020
Due to the extra-ordinary measures taken by the Government and the University, the procedures and protocols for the course have been changed. The University, the School and I, thank you for working with me to coordinate the course at this difficult time.

Please read the following very carefully. ***Each one is responsible to ensure that they do their share***.

The remainder of the course will proceed as follows:
1. To avoid "large gatherings", ***there will be NO more lectures for the course***.
2. The material about Non-parametric estimation, Parzen Windows and Nearest Neighbor methods can be learned from the YouTube video:
   https://www.youtube.com/watch?v=esoVuEG-X1I.
   I will go through this in a separate short talk that I will upload.
3. The material about Perceptron methods can be learned from the YouTube video:
   https://www.youtube.com/watch?v=eJr-M-_qjC4.
   I will go through this also in a separate short talk that I will upload.
4. After these two topics, everything that you need will be covered. You can read about the Ho-Kashyap algorithm at
   http://www.csd.uwo.ca/~olga/Courses/CS434a_541a/Lecture10.pdf.
5. Most questions sent to me by e-mail will be answered in the News file. I myself will not be working via Skype, but will respond, wherever possible, by e-mail, and if necessary, by phone.
6. I will let you know when Assignment 3 has been graded and you can pick it up. Then, we can speak about the submission of Assignment 4 and the Final Project.
7. I thank you again for your cooperation.

March 16, 2020
1. Based on the above news items, I have uploaded a video talk "5107FinalTalk_2020.mp4". It goes through the material in the above tutorials briefly. By listening to my talk and going through the videos, you should be able to do the project.
2. You can find some notes on Fisher's Linear Discriminant at
   https://www.csd.uwo.ca/~olga/Courses/CS434a_541a/Lecture8.pdf.
3. Please remember that you do ***not*** have to understand any mathematical details about the topics. You should just know how the methods can be implemented.
4. I have also uploaded the results of using the Parzen Windows method on unimodal and bimodal distributions.
5. I sincerely hope that everything is clear. Please do the best that you can. I will also be considerate when we grade the project.

March 18, 2020

Here are some further clarification for Assignment #4:

1. For (c), you can plot the learned univariate distributions by sampling at regular intervals, computing the Parzen density estimate, f(x), at each location, x, and plotting the result. This can be done for each of the 3 dimensions and for each class separately.
2. Continuing with (c), to find the sample mean and variance for the learned distribution, use the data accumulated while doing Step 1, above. We can treat all the calculations as i.i.d. samples of a random variable, X, having the learned distribution. The estimated mean is then the expectation, or just the sum of x * f(x) over all samples. You must be sure to first normalize f so that the values of f(x) sums to 1. Similarly, for the estimated univariate variances, you can compute the expectation of $(X - E(X))^2$. The covariance matrix will have these variances along the diagonal, and zeros off the diagonal.
3. An alternative approach is to use a 3-dimensional multivariate Gaussian for the Parzen window. Then, you would use a *multi-dimensional kernel* at every data point to estimate the distribution for any point in the space.
4. For (e), you can classify a test point using the Parzen scheme by first computing the density estimate at the test points using the training points from Class 1. Then repeat this using training points from Class 2. Choose the larger of the two to classify the sample. As there are three dimensions, you can repeat this three times and use a majority vote.

March 23, 2020 (Post A)

The information in Point 2 of the March 18th note regarding the computation of the covariance matrix is not perfect from a multi-dimensional perspective. Here is an updated description.

Consider one class of sample points at a time. There are 9 covariances to work out, 1 for each pair of dimensions. This formula can be used

cov(x,y) = E[(X - E[X]) (Y - E[Y])] = SUM { (xi - E[X]) (yi - E[Y]) * p(xi, yi) },

where X and Y are a pair of dimensions. E[X] has already been computed for each dimension. To compute the covariance, you need to find the joint probability density estimate over pairs of dimensions for each sample point. We can assume that the distributions in each dimension are independent of one another. So, p(xi, yi) = p(xi) * p(yi). You can estimate these probabilities using the Parzen Window scheme once more. To find the mean, you previously computed Parzen density estimates for evenly spaced points. This time, you need to compute Parzen density estimates for each point of the sample. That is, for each point in the sample, you compute the density estimate using all other points in the sample. This gives you p(xi), p(yi), and p(zi). Don't forget to normalize the joint probabilities when computing each expectation for the 9 components of the covariance matrix.

IMPORTANT: *Do the best you can and we will give you credit. You will not be penalized if you made errors in some fine details. Thank you for working together with me during this hectic time*.

March 23, 2020 (Post B)

I have given an extension for Assignment 4. The hard copy of the assignment should be in the box in front of my office by NOON, on Monday March 30, 2020. All the best and please stay healthy.

March 27, 2020

Due to the COVID-19 restrictions, we have made the following changes:

1. Please do NOT drop off the hard copy of anything in the box in front of my office.
2. Please submit Assignment 4 electronically to Omar Ghaleb (omaramin.1992@gmail.com) by NOON, on Monday March 30, 2020. Some of you have already sent it to me. Since I am overwhelmed with mails etc., would you please send it again to Omar?
3. I will post the marks from the first three assignments (without names) on the course website. I will do it soon for Assignments 1, 2 and 3, and then after Assignment 4 has been graded.
4. Please also submit the Final Project to Omar by Noon on the due date.

April 7, 2020

1. The marks of Assignment 4 will be posted by Friday, April 10, 2020 (hopefully). Sorry for the delay, but things are like that during these COVID-19 days. The TA is working hard!
2. From what I understand, everything is working well with the Final Project. Let us see if we can finish it without an extension. ***Then we are finally DONE***!!!

April 13, 2020

1. There have been some questions about whether one needs to test the kNN, Ho-Kashyap and Fisher's discriminant for diagonalized data. The answer is yes.
2. Use the MLE *estimate* of the covariance matrices of the two classes to diagonalize the data. Then one can obtain modified (diagonalized) data. You can use the kNN, Ho-Kashyap and Fisher's discriminant on this new set of points, just as one did for the original points.
3. You will not be able to draw the discriminant function for the kNN. But you can do the best you can for the other classifiers.
4. I will be lenient in grading it.