Fall 2021

COSC522 – Homework #2 (Due 09/16)

Data set used in this homework:

X	Y	Class
0.8	1.2	1
0.9	1.4	1
1.2	1.4	1
1.1	1.5	1
0.8	1.1	2
0.6	1	2
0.65	1.1	2
0.75	0.9	2

1) (45) Mahalanobis distance vs. Euclidean distance.

   a. (10) **Manually** calculate the mean and covariance of the two classes of training samples. You can use calculator for intermediate calculations. However, you need to show details.

   b. (10) Assuming Gaussian distribution, based on the means and covariance matrix, plot the contour maps of the two multi-variate Gaussian distributions for the two classes in Python, overlay the contour on the scatter plot of the data samples. Also plot a test sample $x = [0.85\ 1.15]^T$ on the same figure with different color. Which class do you think x should belong to (based only on visual inspection)?

   c. (5) Write the equations to calculate these two distances between the testing sample and the cluster. (Note: ONLY the equation. Also Note: this is distance not squared distance)

   d. (5) Explain intuitively (in no more than three sentences) the differences between the two distances.

e. (15) Use the following example to understand the differences these two distances make in classification. Here, the minimum distance classifier (i.e., Case 1) is used.

   i. (5) Given a test sample $x = [0.85\ 1.15]^T$, calculate the Euclidean distance to the two class means. Based on the distances, which class should x belong to?

   ii. (5) Use the same test sample, calculate the Mahalanobis distance to the two classes. Based on this pair of distances, which class should x belong to?

   iii. (5) Use kNN with k=1 to label the test sample. Show details. Is kNN with k=1 equivalent to the minimum distance classifier (i.e., Case 1 with equal prior)?

2) (20) Plot the 2-D Gaussian as well as the contour map (i.e., projection of this Gaussian on the x-y plane) with the following covariance characteristics (14 pts). From the plots, elaborate on the physical meaning of each element in the covariance matrix (6 pts – you should learn at least 3 things if given an arbitrary covariance matrix). This is not based on the given dataset. (Note: Suggest to generate a 4x2 plot with the left column the 2-D Gaussian and the right column the contour plot.)

   a. The off-diagonal elements are zero and the diagonal elements are equal to each other

   b. The off-diagonal elements are zero and the diagonal elements are not equal to each other

   c. The off-diagonal elements are positive and the diagonal elements are not equal to each other

   d. The off-diagonal elements are negative and the diagonal elements are not equal to each other

3) (25) Perform agglomerative clustering on the following data sets using dmin and dmax. Need to show details on each iteration. Comment on the different (or similar) shapes of clusters resulted by using these two distance metrics. Use L_1 norm (or city block distance) when calculating dmin and dmax.

   (0.8,0), (2,0), (3.1,0), (4.1,0), (5,0)

4) (10) Using maximum likelihood method to derive the equation for mean and variance assuming the pdf (or likelihood) is modeled by 1-D Gaussian.