Probability and Statistics Spring 2022
HW6 Matlab assignment

1. Bivariate normal distribution

If X and Y are two continuous random variables and they jointly form a normal distribution f(x, y). f(x, y) is called a bivariate normal distribution. In fact, more than two random variables can jointly form a multivariate normal distribution. Here, we look at a bivariate normal distribution as an example.

The probability density function of a bivariate normal distribution can be defined as:

$$\mathrm{f(x, y)} = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}}\exp\left[-\frac{z}{2(1-\rho^2)}\right],$$ where *z* is:

$$\mathrm{z} = \frac{(x-\mu_x)^2}{\sigma_x^2} + \frac{(y-\mu_y)^2}{\sigma_y^2} - \frac{2\rho(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y},$$

ρ is the correlation coefficient of X and Y.

1.(a) Now, please plot four bivariate normal distributions with the following parameters. Show each distribution in a 2D plot in a x-y plane with color representing f(x, y). **You are asked to turn in four plots here.** [Hint: You may consider using *imagesc* and *colormap(jet)*.]

Range of simulation:

X = from 0 to 100

Y = from 1000 to 2000

(For X and Y, an increment step of 1 is good enough. You are free to go finer increment steps.)

Parameters:

Distribution 1: $[\mu_x, \sigma_x, \mu_y, \sigma_y, \rho] = [50, 20, 1500, 200, 0]$

Distribution 2: $[\mu_x, \sigma_x, \mu_y, \sigma_y, \rho] = [50, 20, 1500, 200, 0.3]$

Distribution 3: $[\mu_x, \sigma_x, \mu_y, \sigma_y, \rho] = [50, 20, 1500, 200, 0.8]$

Distribution 4: $[\mu_x, \sigma_x, \mu_y, \sigma_y, \rho] = [50, 20, 1500, 200, -0.8]$

1.(b) From 1.(a), please explain the difference in four distributions and the effect of changing ρ.

Reference:

http://mathworld.wolfram.com/BivariateNormalDistribution.html

2. Decision boundaries

In statistics (and machine learning), people often want to separate samples from two populations (i.e., this is called a classification problem). In this problem, you are asked to simulate two bivariate normal distributions and find out the boundary that these two distribution would be equal. These boundaries are called decision boundaries. Points on the decision boundaries have equal values for both probability distributions (e.g., meaning that they are equally probable to belong to either one of the two bivariate normal distributions). As the name 'decision boundary' implies, points on either side of the decision boundaries belong to the closer bivariate normal distribution. Now, please simulate the following two cases and find out decision boundaries between two distributions.

Range of simulation:
X = from 0 to 100
Y = from 1000 to 2000
(For X and Y, an increment step of 1 is good enough. You are free to go finer increment steps.)

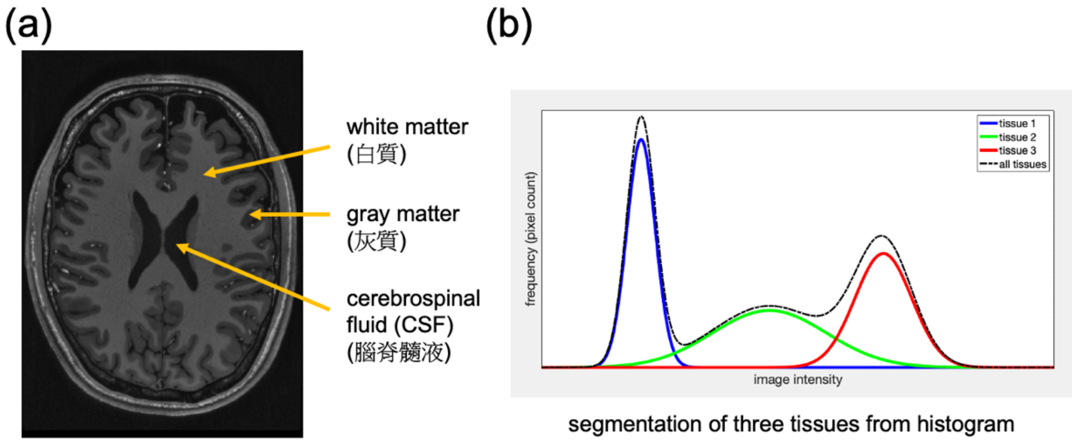
Case 1:
Distribution 1: $[\mu_x, \sigma_x, \mu_y, \sigma_y, \rho] = [25, 30, 1250, 300, 0]$
Distribution 2: $[\mu_x, \sigma_x, \mu_y, \sigma_y, \rho] = [75, 30, 1750, 300, 0]$

Case 2:
Distribution 1: $[\mu_x, \sigma_x, \mu_y, \sigma_y, \rho] = [25, 20, 1250, 200, 0]$
Distribution 2: $[\mu_x, \sigma_x, \mu_y, \sigma_y, \rho] = [75, 30, 1750, 300, 0]$

2.(a) For each case, you are asked to turn in three 2D plots, including two distribution plots and one plot for the decision boundary. [Hint: When finding the decision boundary, it is more convenient for you to check whether the difference of two distributions at a certain (x, y) is lower than *0.005 * max(f(x, y))*.]

2.(b) In addition, you are asked to compare the two decision boundaries and discuss how/why they are (same or different). You are encouraged to experiment with different parameters to see how decision boundaries change accordingly.

3. Image segmentation from intensity histogram



segmentation of three tissues from histogram

In image processing, separation of different objects from images is called image segmentation. A type image segmentation methods separate image objects based on their image intensities. For example, in figure (a), a MRI image of human brain contains three major types of tissues: white matter (bright), gray matter (darker), cerebrospinal fluid (CSF) and image background (darkest). Figure (b) plots an example intensity histogram to illustrate how to separate these three tissues based on image intensity. In figure (b), three Gaussian distributions (blue, green, and red) represent intensity distributions of three types of tissues. The black line plots histogram combined from three Gaussian distributions. Therefore, the three types of tissues can be separated if we can find two proper thresholds to separate three gaussian distributions. In this way, most pixels may be classified to either one of blue, green and red distributions based on their intensity. This problem is designed to bring you some experience of performing image segmentation using your knowledge learned from this course.

3(a) In Matlab, please load 'MRI_brain_14slices.mat', which contains 14 images of human brain scanned by a MRI (Magnetic Resonance Imaging) machine at very high resolution of 250 μm. Please pick an image and display it in gray scale. It should look similar to figure (a). The MRI images are provided by a German research group. If you are interested, please check https://www.nature.com/articles/sdata201732. [Hint: Check commands: *imagesc*, *colormap(gray)*, *set(gca,'dataaspectratio',[1 1 1])*.]

3(b) Please plot an intensity histogram of 14 images. This histogram will be used to determine how to segment different tissues. [Hint: You might want to check command *reshape*. A histogram plot with 200 ~ 300 bins is well enough for you to see distribution of pixel intensity.]

3(c) From the histogram in 3(b), please determine proper intensity ranges for white matter, gray matter, CSF and background, separately. Then, based on intensity ranges you selected, perform image segmentation of three tissues by generating three sets of binary masks (i.e., size of each set should be 880x640x14, 1 for selected pixels, 0 for the non-selected pixels). Display three masks for the image you selected in 3(a). Compare the image you displayed in 3(a) and three masks to get satisfying segmentation results. Adjust intensity ranges manually and redo binary masks if necessary. [Hint: This manual segmentation method is relatively simple. So, don't expect segmentation results to be perfect. Please just make sure that three tissues are separately reasonably well by visually checking their binary masks.]

3(d) Given the intensity ranges you selected in 3(c), please plot three approximated Gaussian distributions of three tissues in the same plot. Please also plot a combined distribution by adding three distributions together. This plot should look similar to figure (b). Compare your combined distribution with the image intensity histogram in 3(b). Comment on whether they have similar shapes.