Question 1: [20 points]

According to theory questions 8 and 9 (1-1 Probability theorem and Bayes' law). We want to add a third class (Sn) for students who have been admitted to the university less than 5 years ago and are currently studying, in addition to other classes (G and nG) mentioned in question 9.

$$p(x,y|G)\sim\mathcal{N}(\mu_G,\Sigma_G) \quad p(x,y|nG)\sim\mathcal{N}(\mu_{nG},\Sigma_{nG}) \quad p(x,y|Sn)\sim\mathcal{N}(\mu_{Sn},\Sigma_{Sn})$$

$$\mu_G=\begin{bmatrix}17\\16\end{bmatrix},\quad \Sigma_G=\begin{bmatrix}2 & 1\\1 & 3\end{bmatrix}$$

$$\mu_{nG}=\begin{bmatrix}13\\15\end{bmatrix},\quad \Sigma_{nG}=\begin{bmatrix}4 & 1\\1 & 5\end{bmatrix}$$

$$\mu_{Sn}=\begin{bmatrix}19\\16\end{bmatrix},\quad \Sigma_{Sn}=\begin{bmatrix}1 & 0.5\\0.5 & 2\end{bmatrix}$$

Your task:

Make a classifier based on a Gaussian Bayes classifier and then:

A) You must plot each class with its boundaries in a 2D space.

B) You must plot Gaussian probabilities for each class in a 3D space.

Question 2: [80 points]

In this section, you have a dataset about people who have taken a COVID-19 blood test. This dataset (covid19.csv) has 34 attributes and one target. You must implement a Bayes and Naïve Bayes (extra credit) classifier to identify people who have tested positive for COVID-19.

A) Separate the dataset into two datasets (Train and Test datasets). It's recommended to split them into 70% for the Train dataset and 30% for the Test dataset respectively (Don't forget to shuffle).

B) Make a Gaussian multi variate PDF for the Training dataset, then implement a Gaussian classifier with uniform probability distribution assumption for all classes (Maximum likelihood estimation). And calculate this classifier accuracy with the Train and the Test dataset.

C) Repeat part B with applying prior knowledge according to the numbers of data in each class (MAP estimation). And compare its accuracy by the Train and the Test dataset with part B classifier.

D) Repeat part B and C with diagonal covariance matrix and compare it with off-diagonal covariance matrix.