**Python/Numpy basics**

Download Iris Dataset from [here](https://archive.ics.uci.edu/ml/datasets/iris). And then run python program to do following in colab.

1. Plot the data points using the first two dimensions (Sepal Length and Sepal Width) .
2. Use three different shapes (triangle, square, circle) to plot datapoints for three different classes. You should use the class information from class label and use them when you decide on shapes (see slide 6 and slide 10)
3. Calculate the mean data point for each class and show them with similar shape with the larger size.
4. Calculate the **centered version** of Sepal Length and Sepal Width
5. Now, plot a line (l) in this plot with line equation. The line is l = span{}. Therefore, the equation of the line is: x1 = - x2 (See slide 10)
6. Now calculate the projection of each data points on the line l (spanned by the vector ). And plot the projected point on the line using the same shape but smaller size. So all smaller shapes would be on the line. (See slide 10)

In 3D:

1. Plot the data points using the first three dimensions (Sepal Length, Sepal Width, Petal Length) .
2. Use three different shapes (triangle, square, circle) to plot datapoints for three different classes. You should use the class information from class label and use them when you decide on shapes (see slide 6 and slide 10)
3. Calculate the mean data point (3D) for each class and show them with similar shape with the larger size.
4. Calculate the **centered version of**  (Sepal Length, Sepal Width, Petal Length) .
5. Now, plot the plane (with yellow plane) spanned by two normal vector( [1 - 2, 1 ]T, [2, 1, 0]T )
6. Now calculate the projection of each data point on the plane (spanned by the vector ( [1 - 2, 1 ]T, [2, 1, 0]T ) And plot the projected point on the plane using the same shape but smaller size. So all smaller shapes would be on the line. (See slide 10)

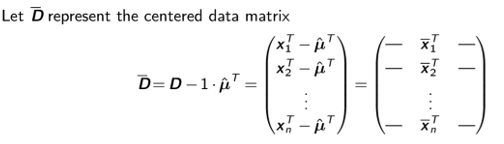
**Numeric Data Analysis**

Download the magic04.data data file from the [UCI ML Repository](https://archive.ics.uci.edu/ml/datasets/MAGIC+Gamma+Telescope). The dataset has 10 real attributes, and the last one is simply the class label, which is categorical, and which you will ignore for this assignment. Assume that attributes are numbered starting from 0.

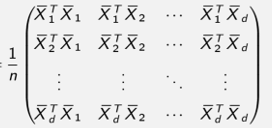
Write a script to answer the following questions.

1. Compute the multivariate mean vector
2. Compute the sample covariance matrix as inner products between the columns of the centered data matrix (see **Eq. (2.38)** in chapter 2).
3. Compute the sample covariance matrix as outer product between the centered data points (see **Eq. (2.39)** in chapter 2)
4. Compute the correlation between Attributes 1 and 2 by computing the cosine of the angle between the centered attribute vectors. Use vector notations shown in class. Plot the scatter plot between these two attributes.

**Analytical Problem:**

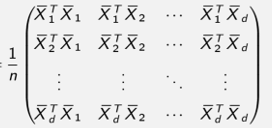
****

**Show that**

1. 

**=**







=



Where ….. are centered attribute vectors. N.B You should use pen and paper

to show the LHS = RHS for a and b here. Use vector notations shown in class. Take the image, make pdf and submit with your ipynb file.