# *Indian Institute of Space Science and Technology*

Department of Avionics

Course Name: Pattern Recognition and Machine Learning (PRML)

# *PRML Programming Assignment 2*

---

1. This assignment is due on: **23/10/2020**
2. This assignment should be submitted either as an **ipython** notebook or a **Latex** compiled document with **codes, results and conclusion**.
3. Use python 3 for implementing.
4. Use **Numpy** (for array processing and linear algebra operations), **Matplotlib** (for Plotting)
5. For reading the data, **pandas** or **python3 inbuilt** file operations can be used.
6. Data files are provided with the link, click on the link text to download the data

---

## Question 1: (Bayesian Decision Theory)

Consider a two-class classification problem where each class is modeled by a 2D Gaussian distribution $G(\mu_1, \Sigma_1)$ $and$ $G(\mu_2, \Sigma_2)$

1. Generate 100,000 samples from each distribution (i.e., 200,000 samples total) using the following parameters (i.e., each sample (x, y) can be thought as a feature vector):

$$\mu_1 = \begin{bmatrix}1\\1\end{bmatrix} \quad \Sigma_1 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix} \qquad \text{and} \qquad \mu_2 = \begin{bmatrix}4\\4\end{bmatrix} \quad \Sigma_2 = \begin{bmatrix}1 & 0\\0 & 1\end{bmatrix}$$

2. Assuming P(ω1) = P(ω2)
   a) Design a Bayes classifier for minimum error.
   b) Plot the Bayes decision boundary **together** with the generated samples to better visualize and interpret the classification results.
   c) Report (i) the number of misclassified samples for each class separately and (ii) the total number of misclassified samples.
   d) Calculate the Chernoff and Battacharya bound on probability of error

3. Repeat 2. For P(ω1) = 0.2 and P(ω2) = 0.8. For comparison purposes, use **exactly the same** 200,000 samples.

## Question 2: (k-NN Classification)

Perform k-NN classification on the iris dataset for classification of the three species task, report the mean accuracy.

For this task first normalize the data (rescale the data) so that, the feature values lie between 0 and 1.

And calculate the mean accuracy across 5-folds of cross validation.

For this problem you can refer to K-NN tutorial for problem 2

## Question 3: (Naïve Bayes Classification)

Implement Naïve Bayes Classification on the iris data set, provided in the last assignment sheet. Take the sepal width, sepal length and petal length features into consideration. (Assume a Gaussian distribution for each feature of each class)

Split the data into training and testing (preferable dichotomy of 70% and 30% respectively) for each class, Report the classification error on testing data.