Exam 3 **CSCE 633**

Due: 11.59pm on November 30, 2021

Instructions for exam submission

Please submit on CANVAS a **single pdf** file containing your solutions.

a) Please write a brief report and **include your code right after each answer**.

b) For each answer, please explain your thought process, results, and observations. Please do not just include your code without justification.

c) Create a **single pdf** and submit it on **CANVAS**. Please do not submit .zip files or colab notebooks.

d) The maximum grade for this exam is **15 points** (out of 100 total for the class).

e) **You can use any available library.**

Question: Clustering countries based on longitude and lattitude

In this exam we will focus on a clustering problem to group countries based on their geographical location.

We will be considering the longitude and latitude of 240 countries, as provided in "data.csv" file on CANVAS. The rows of the file refer to the countries (i.e., samples) of the dataset. The columns denote the country name (column 1), as well as the longitude (column 2) and latitude (column 3) of each country.

(1) (3 points) Data visualization: Plot a 2-D scatter plot of the data. Provide explainations and insights regarding the resulting scatter plot.

(2) (5 points) K-Means Clustering: Apply K-Means clustering to group the samples of the dataset using $K = 2, 3, 4, \ldots, 10$. Provide the color-coded scatter plots for each K, where each color in the scatter plot represents a cluster.

(3) (2 points) Elbow method for K-Means Clustering: Based on the scatter plots from question (2) and using the elbow method, determine and discuss the number of clusters that makes most sense for the data.

(4) (5 points) Clustering with Gaussian Mixture Models: Apply a Gaussian mixture models using 5 Gaussians to group the samples of the dataset. Report the mean and covariance of each Gaussian and briefly discuss these results. How does this clustering result compare to the one obtained by the K-Means algorithm? Please also examine and discuss the log-likelihood of a sample belonging to each Gaussian for 1-2 samples.