Team 15 Checkpoint 1: Project Deliverables & Outcomes

▼ Project Outcomes (High Level)

- To make sure that students, academics, teachers, and individuals at all education levels can easily access and visualize our interactive learning experience.
- Delving into the in-depth theory and intuition behind the Gaussian Mixture Model (GMM) and the Expectation Maximization (EM) Algorithm
- Showing GMM & EM in action in different dimensions of data (1D, 2D, using Principle Component Analysis (PCA) for higher dimensions of data)
- Making the Simulation very interactive where user can manipulate
  - Type of Data (1D, 2D, etc.)
  - Data (User can input or use default data)
  - Perform a step with the GMM using EM algorithm
  - Changing parameters (number of GMM components, number of iterations)

▼ Project Deliverables

Theory Content:

- Explanation of the general EM Algorithm
  - Expectation & Maximization Code Samples
- Explanation of EM Algorithm with GMM

- Comparison with other Clustering Algorithms such as K Means

Visualizations:

- GMM (For every iteration of EM)
  - 1D Data (Animation of gaussian distributions fitting into data)
  - 2D Data (Animation of gaussian ellipses fitting into data)
  - 2D PCA Plots for N Dimensional Data

- Data (Can be N Dimensional)
  - User can either use default datasets
    - Vary number of Clusters in data
    - Gaussian, uniform, or random data
  - Input their own data in the specified format

- User Interaction: (User can manipulate the following hyperparameters and interact with GUI)
  - Number of Components in GMM
  - Number of iterations
  - Play/Pause

▼ By Checkpoint 2

High Level: Making basic web app using React, Dart, Python/Javascript, Manim (Mathematical Animation Engine)

- Start adding the theory content of GMM and EM Algorithm to the webpage
- Implementing EM on GMMs in dart language, python/javascript and visualizing it in Manim for at least 1D data

▼ By Checkpoint 3

High Level: Completing the website as a blog explaining theory of GMM along with interactive visual lessons

- Finishing theory content with code samples and explanation for each part of the algorithm
- Creating visualizations of higher dimensional data (clustering) of GMMs in EM using PCA for dimensionality reduction
- Adding functionality for different datasets and with user given datasets