Helpful details:

The **M-step** updates are given as:

- Priors:

$$\pi_k^{new} = \frac{\sum_{i=1}^N \gamma(z_{ik})}{N}$$

- Cluster means:

$$\mu_k^{new} = \frac{\sum_{i=1}^N \gamma(z_{ik})x_i}{\sum_{i=1}^N \gamma(z_{ik})}$$

- Shared covariance:

$$\Sigma^{new} = \frac{\sum_{i=1}^N \sum_{k=1}^K \gamma(z_{ik})(x_i - \mu_k^{new})(x_i - \mu_k^{new})^T}{\sum_{i=1}^N \gamma(z_{ik})}$$

Programming

4. (**55 points**) In this programming exercise you will implement principal component analysis (PCA) and K-means algorithms. The dataset you will use for this exercise is provided in `data.npy`. All instructions and skeleton code can be found in `hw2.ipynb`.

Submission

- **Things to submit:**

  1. hw2_written.pdf: a document containing all your answers for Problems 1, 2, & 3.

  2. `hw2.ipynb`: a Jupyter notebook containing solutions for Problem 4. Use the skeleton code and fill in the missing parts. Make sure to answer the written questions using markdown.

- **Submit**: All material must be submitted electronically via Canvas. Please **Do NOT** submit `zip` files. Instead, submit two separate files, hw2_written.pdf and `hw2.ipynb`.