

Homework 5 Simulation Exercises

Exercise 1

We are going to use PCA in order to reduce the parameters in images. The dataset contains 213 images belonging to 6 different expressions (Happy, Fear, Angry, Disgust, Surprise, and Sad).

Part A: Plot the eigen values achieved from employing PCA on the dataset. Propose a method which can help determine the proper number of components with regard to total variance. Explain how choosing many components is a vain attempt.

Part B: Display the first four and the last four eigen vectors and explain their differences.

Part C: Plot the eigen values from employing LDA on the dataset. Explain how LDA and PCA differ fundamentally.

Part D: Plot the trace value of the separability matrix ($S_W^{-1}.S_B$). Explain the relation between the trace value and the number of components chosen.

Exercise 2

GMM is a method based on the concept of expectation maximization, and it is used for density estimation and clustering. Use the dataset containing images from Chelsea and Manchester United teams. Consider blue and red colors of each pixel as features.

Part A: set K (number of components) to 2. Employ the GMM algorithm on both classes to estimate their distributions. Display two separate charts containing the contours and means of different classes.