CS-736 Assignment:
Kernel Methods, Statistical Shape Analysis

Instructor: Suyash P. Awate

Please read, carefully, the submission instructions.
Items with **0 points** are a necessary part of the assignment, without which the assignment won't be graded.

1. **(20 points) Shape Analysis on Human Hand Shapes.**

   Get the dataset of shapes from "hands2D.mat" that contains 2D hand-outline shapes represented as pointsets.

   Repeat all the analysis in the first question.

   (a) (0 points)
   Show a plot of the initial pointsets, as given in the dataset. Randomize the color for each pointset, to show all 2D pointsets in one graph.

   (b) (6 points = 3 + 3)
   Show plots of computed shape mean, together with all the aligned pointsets. Show two sets of results, one relying on Code11 and another relying on Code22.

   (c) (6 points = 3 + 3)
   Show plots of the variances (eigenvalues; sorted) along each of the top 3 principal modes of shape variation. Show two sets of results, one relying on Code11 and another relying on Code22.

   (d) (8 points = 4 + 4)
   Show plots of the computed shape mean, all aligned pointsets, as well as two other pointsets depicting the top 3 principal modes of shape variation around the mean ($\pm$ 2–3 standard deviations around the mean). Show two sets of results, one relying on Code11 and another relying on Code22.

2. **(30 points) Shape Analysis on Human Cardiac Shapes.**

   Get the dataset of shapes from “anatomicalSegmentations.zip” that contains 2D images of segmentations of a ring-like shape in the human cardiac anatomy. Assume all segmentations correspond to images that are consistently acquired in similar orientations/poses (like, e.g., faces acquired consistently where top of head is towards top of image, chin is towards bottom of image, etc.), but the images (segmentations or their undelrying medical images) aren’t explicitly registered.

   (a) (10 points)
   For each image, carefully generate a pointset corresponding to the inner and outer boundaries (i.e., the ring). Using these generated pointsets, repeat all the analysis in the first question.

   (b) (0 points)
   Show a plot of the initial pointsets, as given in the dataset. Randomize the color for each pointset, to show all 2D pointsets in one graph.

   (c) (6 points = 3 + 3)
   Show plots of computed shape mean, together with all the aligned pointsets. Show two sets of results, one relying on Code11 and another relying on Code22.

   (d) (6 points = 3 + 3)
   Show plots of the variances (eigenvalues; sorted) along each of the top 3 principal modes of shape variation. Show two sets of results, one relying on Code11 and another relying on Code22.

   (e) (8 points = 4 + 4)
   Show plots of the computed shape mean, all aligned pointsets, as well as two other pointsets depicting the top 3 principal modes of shape variation around the mean ($\pm$ 2–3 standard deviations around the mean). Show two sets of results, one relying on Code11 and another relying on Code22.

3. **(30 points) Robust Shape Mean.**

Get the dataset of shapes from "robustShapeMean2D.mat" that contains 2D elliptical shapes represented as pointsets.

Consider a problem of finding the mean shape, where the given (2D) shapes are represented as pointsets of cardinality N. You have to use standard Procrustes distance to measure dissimilarity between shapes. You are told that the shape dataset has some outliers and you want your clustering to be less sensitive to the presence of outliers in the dataset.

(a) (15 points: 5 + 5 + 5)

Design an optimization problem to estimate the mean shape by penalizing the sum of squared Procrustes distances between the mean shape and the individual shapes.

For the given dataset, implement a function to compute this mean by solving this optimization problem.

Display the original pointsets. Display the estimated mean, and all original pointsets aligned to the mean.

(b) (15 points: 5 + 5 + 5)

Design an optimization problem to estimate the mean shape by penalizing the sum of Procrustes distances between the mean shape and the individual shapes.

For the given dataset, implement a function to compute this mean by solving this optimization problem.

Display the original pointsets. Display the estimated mean, and all original pointsets aligned to the mean.

4. **(40 points) Kernel PCA to Model Variation in Object Segmentations.**

   Get the dataset of segmentations from "anatomicalSegmentations.zip" and "anatomicalSegmentationsDistorted.zip" that contains 2D images of segmentations of a ring-like shape in the human cardiac anatomy.

   To reduce computational complexity in the tasks described next, you may resize the images to size 64×64 pixels, ensuring the pixel values continue to lie within $[0, 1]$.

   (a) (10 points: 2 + 2 + 2 + 2 + 2)

   Implement a function to perform PCA on the (vectorized) segmentation images.

   Show (i) the eigen spectrum, (ii) the mean image, and (iii) the first 2 modes of variation around the mean (as images).

   (b) (15 points: 5 + 5 + 5)

   Using a Gaussian kernel between the (vectorized) segmentation images, implement a function to perform kernel PCA.

   Show (i) the eigen spectrum and (ii) an estimate of the segmentation image that the kernel would (implicitly) map to a point closest to the mean in RKHS; such an estimate is called the *pre-"image"* associated with the kernel.

   (c) (15 points: 3 + 12)

   For each segmentation in "anatomicalSegmentationsDistorted.zip", implement functions to aim to reduce the distortions by projecting the segmentation images onto the first 3 modes of variation in:

   (i) PCA, and display the projected segmentation images, and

   (ii) kernel PCA; here, the projections will be in RKHS; for each of the aforementioned projections in RKHS, implement a function to compute its *pre-"image"* as a segmentation image and display it.