Question1
In this question dataset is a list containing of 70,000 28*28 images each image is resized to 1*784 matrix for applying pca and dataset is split into train_data,test_data

i)from train_data each data point is picked up and a matrix X is constructed with the columns as data points with size 784*60,000 matrix.every datapoint of matrix X is centered and normalized. Then the covariance matrix C of the matrix X is computed.
For the covariance matrix C eigenvalues and eigenvectors have been computed.computed eigenvalues represent the variances and corresponding eigenvectors are the required principal components.eigen values are sorted in the decreasing order accordingly eigenvectors are sorted.
All the eigenvalues corresponding to their principal components are divided by the sum of eigenvalues to find the explained variance associated with each principal components.these acquired principal components are of the form 1*784 these are resized to 28* 28 matrices and the images of the principal components are plotted accordingly.

ii)here i have taken 5 percentage values 25,50,75,95,98.Every time I pick and sum the top eigenvalues which is equal to the selected percentage of total sum of eigenvalues. According to the selected top eigenvalues eigenvectors were chosen and then with the chosen eigenvectors a data point is chosen and projected it on the top selected eigenvectors and summed up all the projections and looked at the images i found that that images which are formed with top 154 principal components(95%) and top 331 (98%)principal components represents the information stored in the image better than images formed with smaller number of principal components.we also can observe information given by images formed by top 154 principal components and top 331 principal components are almost same.hence these top 154 principal components can be used to reconstruct the dataset.hence i will pick 154 as a dimension that can be used for a downstream task.

iii)here in this question i had taken only 1000 points from the dataset and the kernel matrices were computed by using these data points and then computed kernel matrix is centered.eigen values and the eigenvectors of the kernel were found.top 2 eigenvectors and eigenvalues were selected and dataset was projected onto the top 2 eigenvectors and the projections were plotted one against x and one against y.

iv)from looking at the obtained graphs I think the polynomial kernel is the best because projections are something that should store different information for each datapoint in the dataset. From the graph we can observe that the polynomial kernel is better than the radial basis function.
In the case of radial basis kernel since the dataset is too large and sigma is small kernel elements converges to zeros hence so the projections store same information of almost all the data points in the dataset so that the projections is same for most of the data points in the dataset. Hence I would say in the case of larger data points(in terms of magnitude) or data points of larger size it is better to prefer a polynomial kernel over a radial basis kernel with smaller sigmas.

Question 2

i)here i have selected two random integers in range of 1 to 1000 and corresponding to these numbers as indices from dataset two means were picked according to the means each data point is allocated with indicators according to their nearest means and again their means were computed according to assigned previous indicators and again datapoints were assigned their corresponding indicators by nearest of new means and so on in each iteration an objective function is calculated which is the sum of distance of each datapoint to its nearest means and the algorithm will be terminated once the objective function in previous iteration and current iteration are equal.and once the algorithm terminated the data points are plotted on a graph where for each datapoint an indicator is assigned and from the assigned indicator a corresponding color is given to the each datapoint and the data points were plotted.

ii)similarly as above i have taken randomly k means and applied the k means algorithm and ended up with the convergence and the data got clustered and from the clustered data dataset means were obtained and from the obtained means a perpendicular bisector of each 2 means available were taken and plotted and also the clustered data is also plotted on the same graph with each cluster a different color was given.

iii)after choosing a required kernel eigenvalues and eigenvectors were obtained and the top 2 eigenvectors were selected as corresponding to the top eigenvalues and these eigenvectors were put in a column of 1000*2 H matrix.this H matrix is normalized and now each row in the H matrix is a new datapoint and now each data point k means algorithm with randomly initializing k means was applied and at the end a set of indicators is given by k means algorithm.these set of indicator were assigned to the initial data set and correspondingly each datapoint was clustered with the respective indicator colors.
here i had tried the polynomial kernel and radial basis kernel to do the spectral clustering.in polynomial kernel i had tried varying the degree from 1 to 10 but unable to get a useful result but in radial basis kernel also i have tried varying sigma from 0.1 to 1 and 1 to 10 but no useful result when i have taken sigma as 0.08 the data got clustered such that ellipse and parabola shape were individually clustered..

iv)in this question the corresponding index of maximum element in each row of H as indicators and these indicators were assigned to the initial data set and correspondingly clustered and the respective colors were given and plotted.here it thus somewhat cluster the ellipse and parabola data into different sets but equally there is some error like dominating color of data points in parabola were also present in ellipse shape and vice versa.this type of clustering is not as good as spectral clustering.

This shows that there is correspondence between the index of the maximum element in each row of H as indicators and the indicators given by k means on H.at some value of sigma this might become exactly equal and cluster the ellipse and parabola shape into different clusters.