

Final Projects and Presentations

Project Big Data

Objective: Students should analyse a dataset, justify the use of a specific statistical tool (PCA or EM) and present some results.

Previous to presentation.

1-Choose a dataset from the Virtual Campus. The available datasets for the PCA and EM analysis are in the Virtual Campus. You can also consider other datasets.

Presentation (consider some slides using Powerpoint, Prezi or Beamer latex)

2-You should present your dataset (use tables, graphic representations). Describe the type of variables, the number of individuals etc.

3-In the case of a dataset for the PCA show the X matrix, justify the use of Y matrix.

4-For the EM algorithm display the dataset in an image.

5-Briefly, introduce these two techniques

6-Justify the reason for choosing these statistical techniques. For instance, for the PCA explain why you want to use this technique for this dataset. For instance, we want to reduce the dimensionality of the data set because.... Justify the use of this technique (show the R matrix)

7-For the EM algorithm, explain why you want to use this statistical technique. We want to classify the pixels values etc.

8-Explain the necessary steps to analyse the datasets using these two techniques. Show some results, figures etc from these steps.

9-For instance, for the PCA explain why you need the eigenvalues, eigenvectors of the R matrix, shows these results, use the Scree plot, Bivariate plot, Correlation circle etc.

10-For the EM algorithm, explain the two steps involved in this analysis, shows how the iteration process changes in terms of the given initial values for the estimated parameters. Shows the final classification of the pixels values after the EM algorithm. The initial parameter values affects this final classification, show some examples.

11-Explain your results in terms of your datasets

12-Finish your presentation with some final remarks about your analysis.

Remember that you should present two presentations (10 minutes each), one for the PCA and another one for the EM algorithm.