CS5101 Machine Learning

August-December 2021

Programming Assignment - 7

Topic: Principal Component Analysis

Follow the instructions carefully before attempting:

- For implementing the PCA method, you are allowed to use ONLY those python libraries (libraries means numpy, sklearn etc.) which have been taught till now in the ML course.
- You must submit your code in a single python .ipynb notebook with naming format as follows: Firstname_Lastname_assignment7.ipynb
- For each question, create a separate text block containing the question followed by a code block containing the solution.
- Follow each and every instruction given in each question carefully.
- Your code must be properly commented explaining each step clearly.
- If any of the above instructions are not followed, penalty will be there for the same
- Your code and answers will be checked for plagiarism and if found plagiarised, zero marks will be provided for assignment 7. So make sure you actually code and solve the questions rather than noting down the answers.

Task-1:

Generate data samples that are randomly distributed on 3 concentric spheres in 3D space with center at (1,1,1) and radius with 1, 2 and 3 respectively. Plot a 3D scatterplot of them using 3 different colors for the points on 3 different spheres.[1 mark]

Task-2:

Find the 2D presentation of these samples using Linear PCA, PCA with a polynomial kernel with degree 5, and PCA with Gaussian kernel. Please select the width of the Gaussian kernel appropriately so that samples from 3 classes become well separated from each other in their 2D projection. [3 marks]

Task-3:

The data file has 600 observations with 8 attributes. The psychological attributes as: locus_of_control, self_concept, motivation; academic attributes as: read, write, math, science; and one gender attribute. Split the data file into two sets of variables i.e. 'U' and 'V', with 'U' = {set of psychological attributes} and 'V' = {academic attributes + gender}. Perform canonical correlation analysis with:

i. n_components = 2

ii. n_components = 3

where, n_components = number of components to keep, and report the results. [1 mark]