Assignment #4

CE 764 Hydroinformatics

Dept of Civil Engineering, IIT Bombay.

Fall 2017.

Maximum marks: 10. Due date: 11.55 pm, Nov. 8, 2017.

Email to: peterck05@gmail.com, denzilroy@gmail.com with a copy to arpita567@gmail.com

1. Annual average tropical Pacific Ocean sea surface temperature (SST) data is given to you (pacific_sst_data.txt) for latitudes 30.5S to 28.5N, longitudes 109.5 deg to 288.5 (equivalent to 109.5E to 71.5W), with 1 degree increments along both latitude and longitude, for 29 years (1965-1993). Arrange it in a data matrix where number of columns = number of grid points (10,800), and number of rows = number of years (29). Carry out principal component analysis on this data matrix using Matlab.
   i) How much percentage of variance is explained by the first principal component (PC)?
   ii) Plot the first empirical orthogonal function (EOF) spatially. You can use either gridded or contour plots. What do you think this pattern represents?
   iii) Plot the first PC versus time.
   iv) Plot the graph (bar chart) of cumulative percent variance explained (PVE) versus the number of PCs. How many PCs are enough to explain 90% of the variance in the original data?
   v) Plot all the transformed data in the 2D space with the first two EOFs as coordinates. Comment on this plot.

2. Consider the first five PCs from Q1. You are to carry out fuzzy c-means clustering on these five PCs which will further serve as input to regression between their memberships and Indian Summer Monsoon Rainfall (all-India average rainfall during JJAS). This means that your ultimate goal is to predict ISMR from Pacific Ocean SSTs.
   i) Carry out a cluster validity test based on FPI to find out the optimal number of clusters and fuzzification value. Take standard values of FPI to arrive at the optimal values.
   ii) Carry out fuzzy c-means clustering with the optimal values obtained from Q2.i) and compute the membership/partition matrix.

3. Monthly average data for precipitation (ppt), convective precipitation (conv_ppt), sea level pressure (SLP), surface potential temperature (Theta) and outgoing long wave radiation (OLR) from January 1979 to December 2013 are provided (datacca.xlsx) for a particular location (centered at 80°N and 170°E). Arrange the data in such a way that the number of rows = number of observations and the number of columns = number of parameters. Let Y = {ppt, conv_ppt} and X = {SLP, Theta, OLR}.
   i) Carry out canonical correlation analysis on X and Y.
   ii) What is the magnitude of the first and last canonical correlations?
   iii) Plot the first canonical variate pair. Which parameter(s) might have a positive influence on the linear combination of ppt and conv_ppt? Why? Provide physical reasons.

**********end**********