Signal & Data Analysis in Neuroscience
Bar Ilan, Spring 2020 Due: June 18 10:00 AM.

Assignment 8

1) PCA Matlab question

The data file 'dataPCA.mat' contains a matrix called 'lfp'. Each row of the matrix represents LFP recording of one trial, using a sampling rate of 4KHz.
Calculate the PCA of the LFP (using your own algorithm, not matlab's built-in PCA functions):
a) Plot the first 2 principal components.
b) Plot histograms of the projections of each trial onto each of the chosen components.
Also plot a scatter-plot of the projections on the 2 components plain. Explain the graphs referring to classification and representation of the data.
c) What is the percentage of the variance of the data that is explained by the 1st component? By the 2nd component? By both components combined?
What is the percentage of the variance that remains unexplained by the 2 components?
d) Find the vector (dimension) with the highest variance before PCA, and calculate its entropy. Calculate the entropy of data on the first eigen vector. Explain your results.

2) PCA

The MEG signal recorded by two Squids (SA & SB) is distributed Normally N(μ=0,σ=1). During three experiments the following covariance matrices were calculated
(all higher moments are 0):

$$Cov_1(SA,SB)=\begin{pmatrix}1 & -1\\ -1 & 1\end{pmatrix} \qquad Cov_2(SA,SB)=\begin{pmatrix}1 & 0.5\\ 0.5 & 1\end{pmatrix} \qquad Cov_3(SA,SB)=\begin{pmatrix}1 & 0\\ 0 & 1\end{pmatrix}$$

a) Sketch sample values of SA vs. SB for the three experiments (demonstrating the joint distribution).
b) Estimate the eigenvalues and eigenvectors for the three experiments.
c) In which experiment do the sensors display the highest redundancy and in which one the lowest? Explain.

3) Clustering – k means

The file “clusterData1.mat” contains 1000 2D points. Assemble these points into five clusters using k-means algorithm, using the starting points: (-1,-1), (1,-1), (1,1), (-1,1), (0,-1).

a) Implement your own version of the k-means algorithm.

b) Plot the clustered points (each cluster in different color).

c) What are the clusters' centers?

4) Clustering - EM

The data file “clusterData2.mat” contains points sampled from a 2d sample space, originating from several normal distributions.

a) Display all data points on the same figure

b) Implement your own version of an EM algorithm to find the parameters for each cluster, and specify the parameters you found (use k=4).

c) On the same figure as a), plot the contour lines of each of the clusters you found.

d) Repeat the procedure with k=3 clusters and plot your results (data + contour lines) on a new plot.

e) Which k is better? Explain.