

Project - 1 (15 Points)

Published on 29.05.2019

To be collected on 30.06.2019

Problem 1: Regression (5 points)

Based on your understanding of nonlinear regression, try to retrieve a smooth curve from the provided noisy time series as shown below.

Hint: the observed noisy signal could be assumed as a smooth nonlinear signal contaminated by an additive Gaussian white noise.

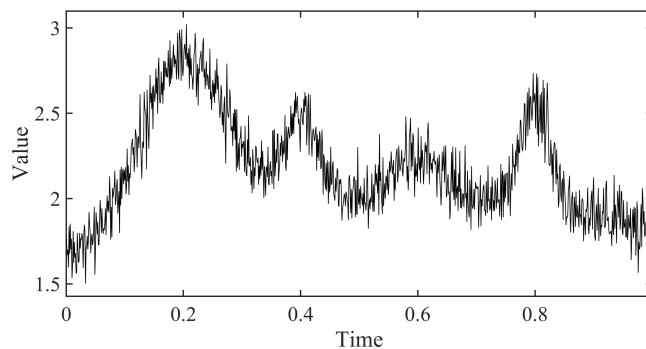


Figure 1: The noisy time series.

Problem 2: Principal Component Analysis (PCA) (5 points)

One interesting application of PCA in facial recognition is to obtain the “Eigen-faces” from well-prepared face dataset. The “well-preparation” implies that all face images in the dataset had been preprocessed, e.g. rescaled, rotated and normalized, such that the images taken from the same person should be similar to each other, e.g. from geometrical point of view. On the other hand, one face image could be considered as a point in a high-dimensional facial space if it was stretched into a long vector, e.g. using “ $x(:)$ ” in the Matlab. Therefore, it is reasonable to assume that any face image taken from the same person could be approximated as a linear combination of most important orthogonal Eigen-faces obtained from the training dataset. For a given test face image, we may easily recognize it by checking the smallest approximation error from some person’s Eigen-faces.

Try to obtain the Eigen-faces from the provided training dataset which contains two groups of cropped face images from extended Yale Face Database B.

Hint: To obtain Eigen-faces from the given training dataset is enough. To check face recognition is not required. More information of the database could be found from here:

<http://vision.ucsd.edu/~iskwak/ExtYaleDatabase/ExtYaleB.html>



Figure 2: A subset of the cropped face images from extended Yale Face Database B. Left: YaleB21, Right: YaleB34.

Problem 3: Nonlinear Clustering (5 points)

Try to construct possible nonlinear kernel for the K-means clustering method such that a clear separation of the provided dataset shown below could be realized.

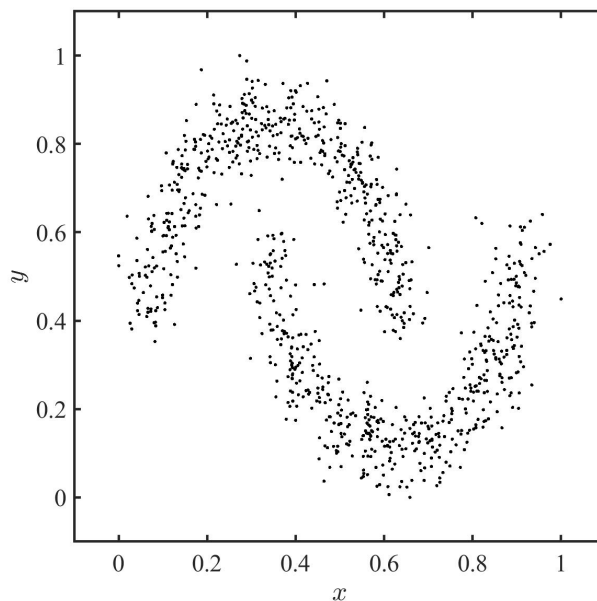


Figure 3: The noisy two moons dataset for nonlinear clustering.

Requirements:

- (1) one can finish the project alone or in group. In the group case, each group member has to write the report individually although a shared group code is allowed;
- (2) since all problems are different, one could prepare the report with multi-sections, each of which contains your statement to the corresponding problem;
- (3) in each section, one could describe and explain following aspects in different sub-sections: a) the considered or assumed model, b) possible ways to solve the problem based on the model, c) the obtained results together with necessary discussions;
- (4) do not forget the citation if some published method has been involved or modified in your report;
- (5) submit the package, including report + codes + results, via E-mail. The deadline may be extended in one more week with justifiable reason.