

DR. BOQIANG HUANG

MATHEMATICAL FOUNDATIONS OF DATA ANALYSIS II — SS 2019

Project - 2 (15 Points)

Published on 16.07.2019

To be collected on 31.08.2019

Problem 1: Support Vector Machine (SVM) (5 points)

Based on your understanding of the SVM with linear or nonlinear kernel, try to realize your own SVM algorithm in order to solve the following classification problem, the famous two spirals under noise.

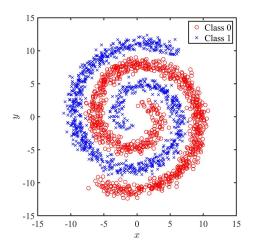


Figure 1: The noisy two spirals as training data set. Red o represents a sample in the class 0, while blue x represents a sample in the class 1.

Problem 2: Independent Component Analysis (ICA) (5 points)

Try to realize your own ICA algorithm and then apply it to the training dataset which consists of many small images with the size 16×16 randomly extracted from given natural images.

Try to explain your obtained independent components (ICs) from different point of views, e.g. feature extraction, data representation with over-complete dictionary.

Reference Paper: J. Hurri, A. Hyvärinen, J. Karhunen, E. Oja, Image feature extraction using independent component analysis. In Proc. NORSIG'96, Espoo, Finland, 1996.

Reference Link: https://research.ics.aalto.fi/ica/imageica/



Figure 2: Four selected natural images from the given image set as an example.

Problem 3: Back Propagation (BP) Neural Network (5 points)

Realize your own BP neural network algorithm, and try to train it for two spirals classification problem. Hint: The training dataset is as the same as the one in the problem 1.

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.