Topics on High-Dimensional Data Analytics
ISYE 8803 - Spring 2019
Homework 3
Due 02/17/2019 11:59pm

Problem 1. Amino Acids (25 points)

Information from five laboratory-made samples is collected. Each sample contains different amounts of tyrosine, tryptophan and phenylanine dissolved in phosphate buffered water. The samples were measured by fluorescence (emission 250-450 nm, excitation 240-300 nm, 1 nm intervals) on a spectrofluorometer. The array to be decomposed is then $5 \times 201 \times 61$, it can be found as X in the file aminoacid.mat. Ideally, these data should be describable with three CP decomposition components, because each individual amino acid gives rank one contribution to the data.

a. (15 points) Implement the alternative least square algorithm to obtain a rank-3 decomposition of the data (write your own algorithm).

b. (5 points) If the decomposition is appropriate, it should have three components and therefore a 5×3 score matrix (first mode loading matrix). Each column in the score matrix should approximately match the concentration of one of the three amino acids, which are held in the 5×3 Y matrix. Present the score matrix and comment on your results.

c. (5 points) Present the emission and excitation loadings of the CP decomposition. Comment on your results.

Problem 2. Cats and birds (50 points)

The goal of this problem is image classification. There are two classes: cats and birds. You are given 28 training images, train1.jpg through train28.jpg. The first 14 images correspond to cats, and the remaining images correspond to birds. The labels for the images can be found in the file train_lab.mat. Your job will be to classify 12 new images, Test1.jpg through Test12.jpg. To do so follow the following steps.

a. (20 points) For each image on the training sample, use Tucker decomposition with $R_1 = 10, R_2 = 10, R_3 = 3$. Use the core tensor of the images to train a random forest with 100 trees. Note that for each image you will need to vectorize the core tensor obtained. You can use the MATLAB function 'TreeBagger´ to build the forest. Predict the label for the images on the test set. Note that you will also need to use Tucker decomposition for the images on the test set. Report the classification error per class. If some of the images are miss-classified, explain why this can be the case.

b. (20 points) Use Sobel operator to detect the edges of each image. Use 150 as threshold. Use the black and white images for classification. For each image on the training sample, use Tucker decomposition with $R_1 = 10, R_2 = 10$. Use the core tensor of the images to train a random forest with 100 trees. Predict the label for the images on the test set. Report the classification error per class. If some of the images are miss-classified, explain why this can be the case.

c. (10 points) Compare the two approaches.

Problem 3. Heat Transfer Process (25 points)

Consider a heat transfer process that follows the following equation:

$$\frac{\partial S(x,y,t)}{\partial t} = \alpha \left(\frac{\partial^2 S}{\partial x^2} + \frac{\partial^2 S}{\partial y^2} \right)$$

where $0 \leq x, y \leq 0.05$ represents the location of each image pixel, α is the thermal diffusivity coefficient, and t is the time frame. The initial and boundary conditions are set such that $S|_{t=1} = 0$ and $S|_{x=0} = S|_{x=0.05} = S|_{y=0} = S|_{y=0.05} = 1$. At each time t, the image is recoded at locations $x = \frac{j}{n+1}, y = \frac{k}{n+1}, j, k = 1, \cdots, n$, resulting in an $n \times n$ matrix. Here we set $n = 21$ and $t = 1, \cdots, 10$, which leads to 10 images of size 21×21, that can be represented as a $21 \times 21 \times 10$ tensor.

The thermal diffusivity coefficient depends on the material being heated. In the dataset heatT.mat, we have tensor 1 corresponding to a heat transfer process in material 1, and tensor 2 corresponding to a heat transfer process in material 2. Additionally, we have a third tensor. The goal of this problem is to determine to which material does tensor 3 correspond.

In order to reach the goal, use CP and Tucker decomposition. Use different ranks for decomposition. Present your conclusion and the steps that lead to it. In particular, which decomposition method gives better results?