**ISyE 8803: Topics on High Dimensional Data Analytics**

**Homework 3**

**Question 1. Tensor Decomposition Reconstructions (15 points)**

**Part 1.** Kruskal tensors are a way of representing tensor decompositions as a weighted sum of outer products.

for each rank of the decomposition, r, and rank of the original tensor, n.

a) Given the following rank-2 CP decomposition:

Write out the calculation of the first outer product

b) Either by hand or in code, calculate:

* the full reconstruction.

**Part 2.** A Tucker decomposition of the same original tensor is:

Compute the reconstruction of the Tucker decomposition.

**Part 3.** The actual original tensor was:

Calculate the MSE for both the CP and Tucker decompositions. Briefly discuss (2-3 sentences should be sufficient) the difference, especially regarding the relative reduction of features for each method.

**Question 2. Multilinear Algebra (20 points)**

Given , = , and , find the vector of regression coefficients by solving the following optimization problem:

Simplify the above expression to an appropriate form before solving the optimization problem.

Hint: .

**Question 3. Image Classification (35 points)**

Dimensionality reduction, feature extraction and selection are crucial parts of high multidimensional data analysis. Consider a set of training samples corresponding to categories/classes, and a set of test data . The challenge is to find appropriate labels for the test data. The classification algorithm can be generally performed in the following steps:

1. Find a set of basis matrices and the corresponding features from the training data The relation of a sample and basis factors can be expressed as:

Where the core tensor representing features of a much lower dimension than the training data . In other words, the core tensor consists of features of in subspace .

1. Perform feature extraction for the test samples using the basis factors found for the training data (using a projected filter).
2. Perform classification by comparing the test features with the training features.



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. There are two classes: cats and 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. Use the training features to train a random forest with 100 trees. Note that you will need to vectorize the training features. You can use the MATLAB function “TreeBagger” to build the forest.

**Part 1.** Read and convert all images into gray scale. Form a third-order tensor using the training data and apply Tucker decomposition with . Predict the labels for the images on the test set. Report the classification error.

**Part 2.** Read all images in RGB format. Form a fourth-order tensor using the training data and apply Tucker decomposition with . Predict the labels for the images on the test set. Report the classification error.

**Question 4. Heat transfer process (30 points)**

Consider a heat transfer process that follows the following equation:

where represents the location of each image pixel, is the thermal diffusivity coefficient, and is the time frame. The initial and boundary conditions are set such that and . At each time , the image is recorded at locations , resulting in an matrix. Here we set and , which leads to 10 images of size 21 × 21, that can be represented as a 21×21×10 tensor.

The thermal diffusivity coefficient depends on the material being heated. In the dataset heatT.mat, we have tensor 1, tensor 2 and tensor 3 corresponding to a heat transfer process in material 1, material 2 and material 3, respectively.

**Part 1.** Try different ranks for CP decomposition and use AIC to choose the optimal one.

**Part 2.** Use CP decomposition to decouple temporal and spatial patterns of the three materials in heat transfer processes. Plot the first 4 spatial and temporal patterns of tensor 1.