Algorithms for Imaging Spectroscopy, Fall 2018
Project 1
Due Date: Wednesday October 3

The goal of this project is to understand dimensionality reduction hyperspectral data. Since dimensionality reduction without a goal is difficult to analyze, you will evaluate the performance of different algorithms by their effects on classification.

You are given a two-class problem of distinguishing two different classes of trees. The spectra, wavelengths, and class labels are all contained in the file

Proj1Data.mat

The steps you should take are:

Remove water bands and bad bands at the low and high wavelengths

For n = 1:NumberOfExperiments

1) Randomly select 75% of the samples of each class for training. The other 25% will be used for testing

2) Perform dimensionality reduction using PCA, MNF, Hierarchical Dimensionality Reduction, and Downsampling by building a transformation on the entire training set (You may have to use MATLAB for this part), generating reduced dimensionality training set T_{rn}.

3) Apply the transformations to the testing set, generating reduced dimensionality testing set T_{en}.

4) Train a Support Vector Machine classifier on T_{rn}
    a. svmc in SciKitLearn
    b. ScoreSVMModel = fitPosterior(SVMModel,X,Y) in Matlab

5) Calculate the percentage of samples from T_{rn} that are correctly classified by the SVM.

6) Calculate the percentage of samples from T_{en} that are correctly classified by the SVM.

7) Perform dimensionality reduction using PCA, MNF, Hierarchical Dimensionality Reduction, and Downsampling by building a transformation for each class from the training data and concatenating the outputs, creating training and testing sets ST_{rn} and ST_{en}.

8) Repeat 4)-6) with ST_{rn} and ST_{en}.

end

You should submit

A 7 page, single-column, double-spaced report with 12-point font with contents:

- Overview of Methods
- Implementation Notes
- Experimental Results (using tables from below)
- Observations

Functions:

[Y, WY] = ReduceDim(X, W, N) with

- Inputs: X: Spectra
  - W:Wavelengths
  - N: Number of reduced dimensions
- Outputs: Y: 4 Sets of reduced dimensionality spectra, 1 for each algorithm
  - WY: Output Wavlengths (if applicable)

Other ***Optional*** Inputs and Outputs that you want

[OutLabelsTrain, OutLabelsTest] = Classify(TrainSet, TestSet, LabelsTrain);

PercCorrect = ScoreClassifer(TrueLabels, OutLabels)

RunAll: A script that uses these functions and

Runs the experiments described above

For each dimensionality reduction method, Prints 2 Named formatted arrays

A NumberOfExperiments+2 x 4 array of Training Correct Percentages (CPs)

- 1st NumberOfExperiments rows: CPs for each experiment and algorithm
- Row NumberOfExperiments+1: mean CPs for each algorithm
- Row NumberOfExperiments: standard deviations of CPs for each algorithm

Similar array for Testing

CPs should be representing as integers by multiplying by 100 and rounding

Means and Standard Deviations can be floats calculated by multiplying by 100

- Do not use more than 1 digit to the right of the decimal point

A powerpoint presentation of not more than 14 slides. You will present this on the due date.

Grading will be based on the following criteria, with each criterion contributing 25% of the grade.

- Readability of report, code, and powerpoint
- Correctness of code
- Reasonableness of results
- Meaningful observations