# Algorithms for Imaging Spectroscopy, Fall 2018
# Project 1
# Due Date: Wednesday October 10

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.

**The Data Set:**

You are given a two-class problem of distinguishing two different classes of trees.

Scatter plots of the 2 classes are shown in Figure 1.

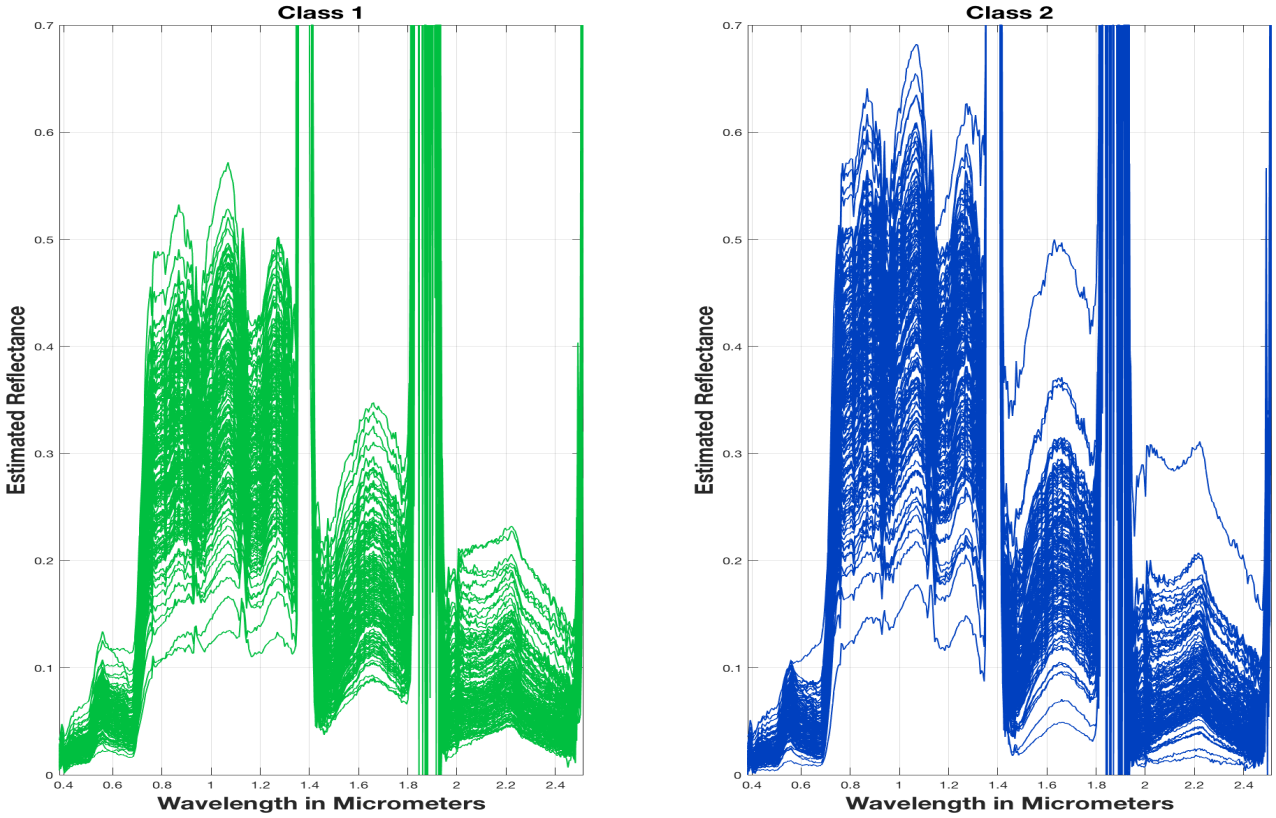



***Figure 1. Scatter Plots of the two classes.***

Plots of the classes as distributions on ($w$, $r$) where $w$ and $r$ are wavelength and reflectance are shown in Figure 2.

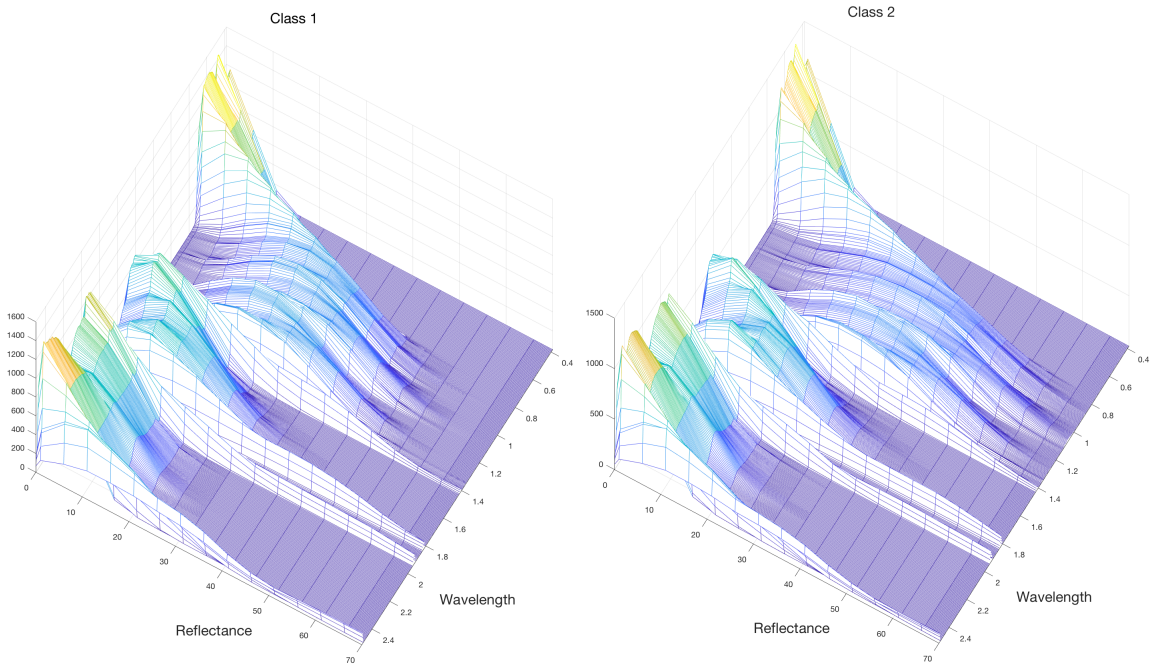



***Figure 2. Distributions of the two classes as functions of wavelength and reflectance.***

The spectra from the classes were extracted several images. The original locations of the spectra in the images are not available. Therefore, two fake images were constructed. The first, called ClassIm, contains all the spectra from both classes. There are 1624 samples in each class. ClassIm is a 16 x 203 x 426 hyperspectral cube for which the first 8 rows are spectra from Class1 and the second 8 rows are spectra from Class2. The mean over all spectra is shown in Figure 3.

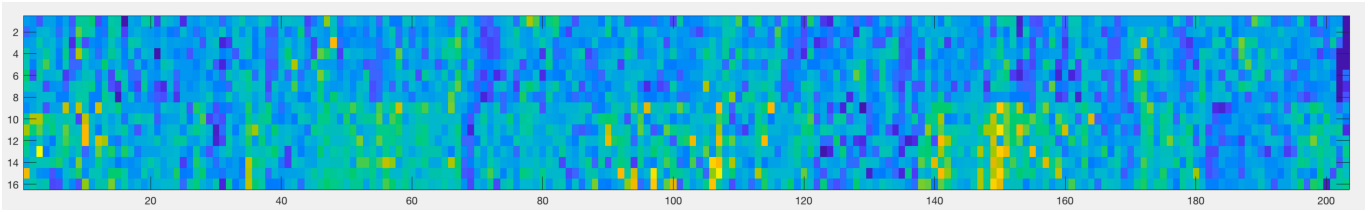


***Figure 3. Mean over all wavelengths of fake image containing all samples from Class1 in rows 1-8 and all samples from Class2 in rows 9-16.***

Another fake image, called EstNoiseFromMe, is a 32 x 406 x 426 image with the spectra from ClassIm placed in every other row and column. You should add noise to this image and then estimate the noise from the image so you can do MNF. It is shown in Figure 4.

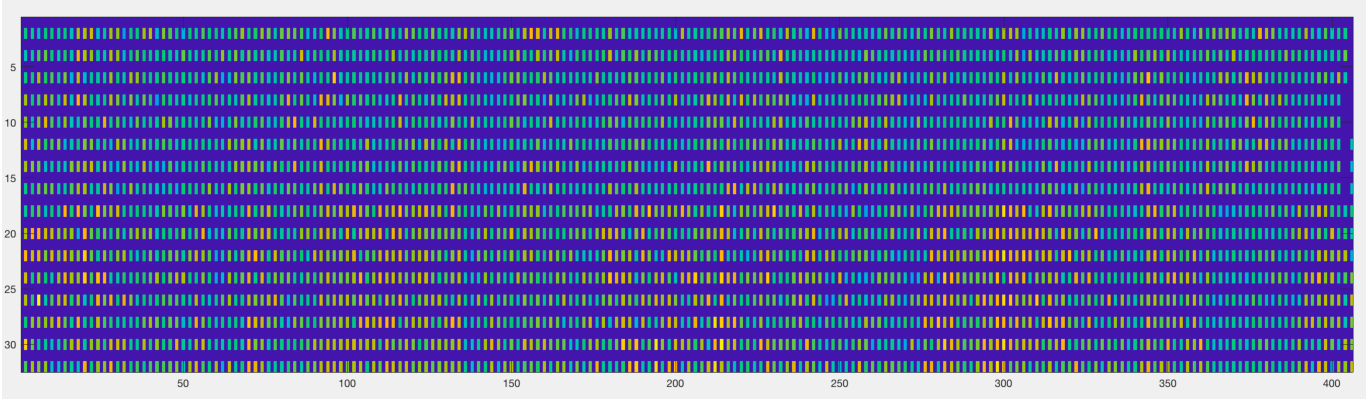


***Figure 4. Mean over wavelengths of a fake image, EstNoiseFromMe, constructed by placing the spectra from ClassIm in every other row & column. Add noise to EstNoiseFromMe and then use it to estimate the noise.***

The data and code are contained in the dropbox folder:
TeachingData/ImagingSpectroscopy/Project1.

The spectra, wavelengths, and class labels are all contained in the file **Proj1Data.mat**. The contents of usage of the other files should be clear but let me know if they are not.

## The Experiments.

The steps you should take are described in the algorithm below. There are some additional investigations to conduct in the context of some of the algorithm steps that are described following the algorithm.

ALGORITHM: Proj1Alg

Remove $H_2O$ bands and noisy bands at the low and high wavelengths.

For n = 1:NumberOfExperiments

1) Randomly select 75% of the samples from each class for training. The other 25% are 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 may wish to use other classifiers but your report should only contain the results using the SVM.

Additional investigations should be investigated outside the context of **Proj1Alg** at first but then should be systematically incorporated into **Proj1Alg**:

Experiment with the effects of adding noise with different covariances.

Experiment with the diagonal loading factor in the MNF program. What happens when it is low? What happens when it is high? Is it necessary?

Experiment with different methods of calculating the KL-divergence and with combing bands in the hierarchical clustering.

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

**Addendum. Understanding the effects of noise.**

For this task, you will investigate the effects of differences in noise estimation using the data in the file:

FakePaviaThings.mat

The file contains the following

```
FakePaviaIm
FakePaviaImSq
PaviaColors
ColorNames
PaviaWavelengths
```

where

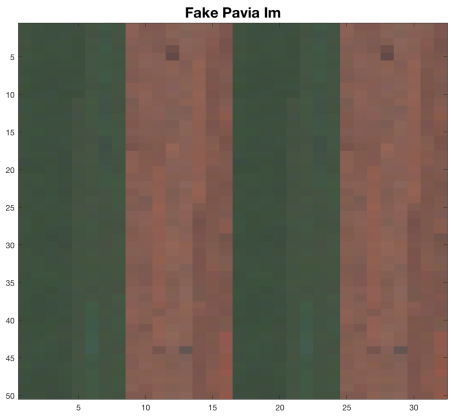
```
PaviaColors = [64, 32, 10] are Red, Green, and Blue indices into of Pavia Data
ColorNames  = {'Red', 'Green', 'Blue'}
```

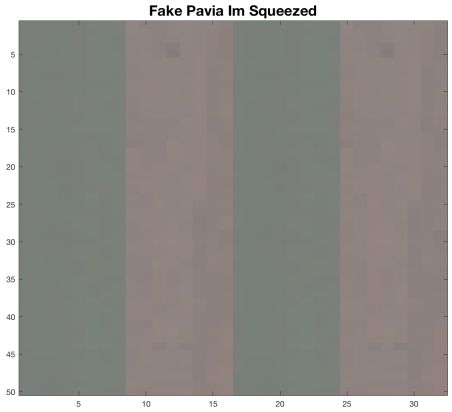
and the images and corresponding spectra are shown in Figure 5.

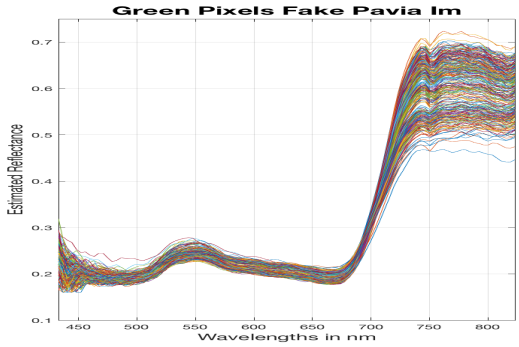
Write a program that

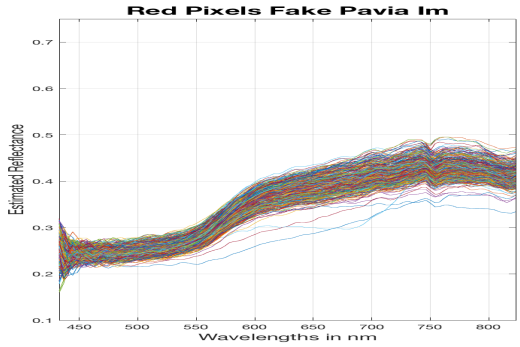
```
Adds Increasing Levels of Noise to FakePaviaIm
Adds Increasing Condition Numbers to Covariances
Applys MNF to the result

Adds Increasing Levels of Noise to FakePaviaImSq
Adds Increasing Condition Numbers to Covariances
Applys MNF to the result
```

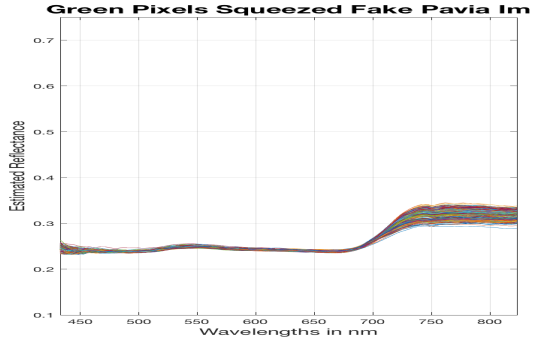
Make observations on the effects of noise. You don't have to run the classification programs again.

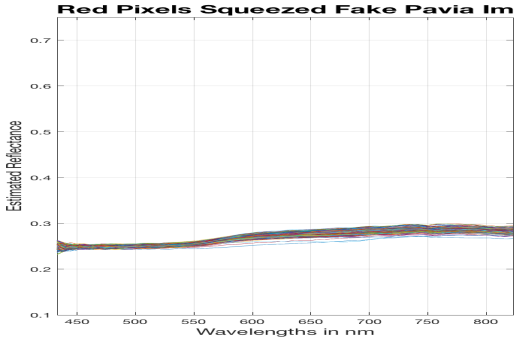

Fake Pavia Im


Fake Pavia Im Squeezed


Green Pixels Fake Pavia Im
Estimated Reflectance
Wavelengths in nm


Red Pixels Fake Pavia Im
Estimated Reflectance
Wavelengths in nm


Green Pixels Squeezed Fake Pavia Im
Estimated Reflectance
Wavelengths in nm


Red Pixels Squeezed Fake Pavia Im
Estimated Reflectance
Wavelengths in nm