Project 2 – Unsupervised Learning – MVA – ENS Paris-Saclay

Instructor: René Vidal
TA: Argyris Kalogeratos

Due Date: 11/24/2017

INSTRUCTIONS: This project is to be done in MATLAB or Phyton and in groups of three students. Please report which student did which part. Each student should also email TA a grade (0-100%=best) for the work of his/her mates.

1. **Clustering Algorithms.** Implement the Spectral Clustering (SC) algorithm (Algorithm 4.5), K-Subspaces algorithm (Algorithm 6.1) and Sparse Subspace Clustering (SSC) algorithm with noisy data (Algorithm 8.5 and 8.6). You can use the MATLAB `kmeans` function (or python) inside SC.

```
function groups = SpectralClustering(Affinity, n)
% Affinity: N by N affinity matrix, where N is the number of points.
% n: number of groups
```

```
function [global_groups, global_obj] = ksubspaces(data, n, d, replicates)
% data: D by N data matrix.
% n: number of subspaces
% d: dimension of subspaces
% replicates: number of restarts
```

```
function groups = SSC(data, n, tau, mu2)
% data: D by N data matrix.
% n: number of clusters
% tau, mu2: parameter
```

   Implement also a function for computing the clustering error. Note that the true labels for 9 points in two subspaces may be `[1 1 1 1 2 2 2 2 2]` while an algorithm may produce `[2 2 2 2 1 1 1 1 1]`, which is a correct clustering. Therefore, you need to evaluate the error for all possible permutations of n letters, where n is the number of groups. When n is large, this can be time consuming. You may use an existing implementation of the Hungarian algorithm to do this efficiently.

```
error = clustering_error(label, groups)
% label: N-dimensional vector with ground truth labels for a dataset with N points
% groups: N-dimensional vector with estimated labels for a dataset with N points
```

2. **Face Clustering.** Apply the three algorithms in part (a) to the face images from the ExtendedYaleB dataset. The data is already in matrix format. The file `ExtendedYaleB.mat` contains two variables `EYALEB_DATA` and `EYALEB_LABEL`. `EYALEB_DATA` is a data matrix containing face images of 38 subjects, each under 64 different illumination conditions. Each column contains a face image of size 48*42, which has been flattened to a vector of size 2016. Use the following command to show a face image:

```
reshape(uint8(EYALEB_DATA(:, 1)), [48, 42]))
```

EYALEB_LABEL is a vector where each entry is the label (ranging from 1 to 38) for the face in the corresponding column of EYALEB_DATA. Report the clustering error for individuals 1-2 for different choices of the parameters σ and K for SC (when constructing a Gaussian affinity with K-NN), number of restarts for K-subspaces, and λ for SSC. Once you have chosen the best parameters for each method, report the clustering errors for individuals 1-2, 1-10, and 1-20, 1-30, 1-38, and analyze the clustering error as a function of the number of groups.

3. **Motion Segmentation.** Apply the algorithms in part (a) to the feature point trajectories from the Hopkins155 dataset. You may use the code below to extract the data matrix.

```
d = dir;
for i = 1:length(d)
    if ( (d(i).isdir == 1) && ¬strcmp(d(i).name,'.') && ¬strcmp(d(i).name,'..') )
        filepath = d(i).name;
        eval(['cd ' filepath]);

        f = dir;
        foundValidData = false;
        for j = 1:length(f)
            if ( ¬isempty(strfind(f(j).name,'_truth.mat')) )
                ind = j;
                foundValidData = true;
                break
            end
        end
        eval(['load ' f(ind).name]);
        cd ..

        if (foundValidData)
            n = max(s);
            N = size(x,2);
            F = size(x,3);
            D = 2*F;
            X = reshape(permute(x(1:2,:,:),[1 3 2]),D,N);
        end
end
```

Report the clustering error for different choices of the parameters σ and K for SC (when constructing a Gaussian affinity with K-NN), number of restarts for K-subspaces, and λ for SSC. Compare the best results.

Submission instructions. Please send an email to the TA with subject `MVA-Supervised-Learning-Project2` and attachment `lastname1-lastname2-lastname3-MVA-Supervised-Learning-Project2.zip` or `lastname1-lastname2-lastname3-MVA-Supervised-Learning-Project2.tar.gz`. The attachment should contain a file called `README`, which contains instructions on how to run your code. The TA will run your scripts to generate the results. Please include a report in PDF format containing figures and tables with your results as well as an explanation of what you did an analysis of the results.