Assignment 1

Files are to be uploaded in

`https://drive.google.com/drive/folders/1UW8518QcegUeq-DhnWS7asTlQVXfzy-r?usp=sharing`

And the following document, *Progress* is to be completed online.

`https://docs.google.com/spreadsheets/d/16xZ9af_kchAAO2F_N4isp01Ggk0fNucT9_F6OD9wE8E/edit?usp=shari`

General questions and comments should be reported on the following document. It is important that a number and a letter be assigned to each question.

It is important that the groups doing projects upload their work as soon as possible to enable reviewers to go through all checking before everyone vote. Each attendant may vote for as many projects and as many reviewers as he wants. It is an indication for me of what you find most interesting and this is most valuable information for me.

1 Assignment for those who are achieving projects

A project is composed of three files. The first file is a `.pdf` document answering to the following questions. Its name should be `project` followed by a number and then by an `a` indicating that this is the first assignment.

- When the data has been collected?

- Where is the area monitored?

- What is it expected to be used for?

- Why is it made publicly available?

- What are the wavelengths collected?

- What are the different classes in the ground truth?

- What is the spatial resolution?

- Has the data been calibrated?

Further information upon the classification are welcome. Which classes are most important to discriminate, which wavelength is expected to be most useful? It could be interesting to change the goal of classification. The publicly available dataset is a material, on which you could try very different tasks. Especially you could have in mind a very specific application.

The second file is a `.mat` containing the data. It should not exceed 10Mo. [1] It should have exactly the same name. It contains the following Matlab objets.

- `X` is a matrix. Each line refers to a pixel of an hyperspectral image. Each column refers to a wavelength. [2] Each component should be a non-negative value.

- `Y` is a column-vector. Each component refers to a pixel. Each component is a integer value indicating the true class to which the pixel belong.

- `metadata` is a structure with supplementary information. It contains the following fields.

  - `size` is the size of the image, it is a row-vector.
  - `features_s` is a structure containing information on the different wavelengths. Each field of this structure has the name of a set of features, and is equal to a row vector indicating the corresponding column numbers.
  - `F` is the number of features.
  - `Void_pixels` is the number of pixels whose class is unknown in the original hyperspectral image.
  - `classes_l` is a cell containing a list of the names of each classes.
  - `i` is a number of a line
  - `j` is a number of a row
  - `x_ij` is a list of intensities associated to each wavelength for the pixel indicated by `i` and `j`.
  - `y_ij` is the class to which the pixel belongs to.
  - `C` is the number of classes.

---

[1]To reduce the size of a dataset, a smaller part of the hyperspectral image could be considered and the intensities associated to each pixel and wavelength could be stored as an integer each being stored with one or two bytes, using for instance `uint8` or `uint16` in Matlab.

[2]In machine learning, it is general practice to call features the columns of a data matrix.

– `X_with_null_pixels` it is the same X matrix but with pixels whose ground truth are equal to zero.

Note that a value of `0` for the ground truth image indicates that there is no adequate ground truth for that pixel. If in the ground truth, there are no pixels belonging to that class, or an extremely small number of pixels, then that class should be removed and the other classes should be renamed accordingly.

The third file is a Matlab script named `PNA_show`. [3] There are two ways of using that script.

- `figure(1); imshow(PNA_show('projectNA','data',3));` makes an image in shades of gray showing the intensity associated to the third wavelength. The title of the figure indicates roughly some indication on the wavelength.

- `figure(1); imshow(PNA_show('projectNA','GT',3));` makes an image in black and white showing in white the pixels belonging to the third class. The title of the figure indicates the name of that class.

Many other tasks could also be implemented as long as they do not contradict these specific syntaxical rules.

All documents used to answer the different questions should be put in a folder that is then zipped into a file named `project` followed by the number and by `.zip`.

2 Assignment for those who are reviewing projects

The goal is to build matlab functions that achieve some basic checks on the data provided along each project. Two files are to be delivered.

The first file is a `.pdf` document. Its name is `reviewer` followed by a number and an `a` indicating that it refers to the first assignment. The first part of this document explains what is tested by each test. The second part explains for each project what has passed and what has failed with precise values showing the problem. The third part is optional, it explains what supplementary information you would request from the projects and how this information could provide more valuable testing.

The second file is a `.m` script having the same name, it runs successively the different functions contained in this file that do the different testings.

Here is a set of checks that could be implemented.

- The saved file has at least the requested variables.

- The size of `X` and `Y` is consistent with the size of the image stored in `metadata`.

- The size of `X` is consistent with the number of features indicated in `metadata`.

- The size of `Y` is consistent with the number of classes indicated in `metadata`.

- values of `X` and `Y` for a specific sample is consistent with the values of both images at a specific pixel coordinate indicated in `metadata`.

Many other checks could also be implemented.

3 Discussion

Your task is first of all to read all projects and check `Progress`. You should write a single `.pdf` document, named `discussionA.pdf` discussing how all projects have undergone this first step, the difficulties that have been overcome and those that remain challenging issues. You should then express your opinion as to whether I should come back on some specific issues. You may also add some specific comments to a specific project on *Discussions* [4] and some specific questions on *Questions*. You are also expected to write in *Questions* the answers to all other questions.

---

[3]N indicates here the number of the project.

[4]Comments should be most respectful as any work needs attention, and regardless of it being possibly wrong, it is going to be useful to get a better understanding. So there can be no shame in being wrong.