



Department of Statistics & Data Science Institute
Columbia University

GR5293 Applied Machine Learning for Image Analysis Assignment #3

Instructor: Xiaofu He

Fall 2020

Goal

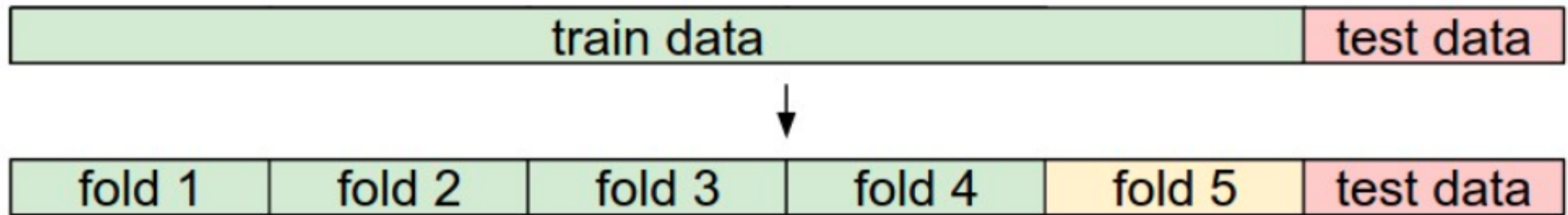
Implement SVM or other machine learning methods for brain state classification (classification accuracy can be based on the whole brain or region of interest from the fMRI data).

Requirement

- a) Download data from coursework. For sub-01/ses-test, load in data and labels.
- b) Preprocessing: realignment (optional); create a brain mask to extract brain regions (i.e., exclude non-brain voxels) using a single threshold and save the mask.
- c) Feature selection: extract brain regions using the brain mask, i.e., exclude background (non-brain voxels) , and/or use dimension reduction (optional) such as PCA.
- d) Classification: SVM or other method
- e) K-fold cross-validation (e.g., K=4, 8...)
- f) Play around the parameters (e.g.,. parameters (brain mask threshold, SVM parameters, number of PCA components (**optional**)). Submit the best accuracy ($\geq 85\%$) and record all parameters.

Repeat **step (a) to (f)** using sub-01/ses-retest data

For example: 5-fold cross-validation



Common data splits. A training and test set is given. The training set is split into folds (for example 5 folds here). The folds 1-4 become the training set. One fold (e.g. fold 5 here in yellow) is denoted as the Validation fold and is used to tune the hyperparameters. Cross-validation goes a step further and iterates over the choice of which fold is the validation fold, separately from 1-5. This would be referred to as 5-fold cross-validation. In the very end once the model is trained and all the best hyperparameters were determined, the model is evaluated a single time on the test data (red).

Extra credit

1. Realign sub-01/ses-test and sub-01/ses-retest together (optional), so that you can train your model using sub-01/ses-test data only and test your model using sub-01/ses-retest. You will receive extra 5% point if you can do so and achieve descent accuracy ($> 70\%$) (It will be challenging)

Submission (Option 1 Matlab)

- One Matlab file, i.e., combine all steps into one script, which must be runnable (we won't debug for you).
 - UNI_Name_Assignment3.m
- A readme file (UNI_Name_Assignment3.README)
 - How to run your code? Any variable needs to be changed before we run your code (e.g., variable for directory)?
 - Briefly discuss the limitation(s). How can you improve it?
 - Describe the results of the experiment in your own words. Compare the results of two approaches (e.g., SVM only vs PCA+SVM) and briefly discuss why one works better/worse than the other.
- DO NOT submit the dataset
- Compress all files and folder into a single compressed file with UNI_Name_Assignment3.zip/.rar as its name

Submission (Option 2 Python)

- One python file (i.e., py), i.e., combine all steps into one script, which must be runnable (we won't debug for you).
 - UNI_Name_Assignment3.py
- A readme file (UNI_Name_Assignment3.README) requirements see previous slide
- DO NOT submit the dataset
- Compress all files and folder into a single compressed file with UNI_Name_Assignment3.zip/.rar as its name

Submit to the coursework, due on
11/18/2020 (11:59PM)

Hints

For Matlab

- use the following SPM functions to read your fMRI data
 - `spm_vol`: Get header information for images
 - `spm_read_vols`: Read in entire image volumes
- e.g.,
- ```
imgHeader = spm_vol(imgFile);
imgData = spm_read_vols(imgHeader);
```
- feel free to use existing toolboxes that simplify the work for you

## For python

- scipy.io for loading .mat data file if you processed the data in Matlab or use a Python package (e.g., nibabel, see an example at <https://nipy.org/nibabel/gettingstarted.html>) to load nifty data
- feel free to use other packages to make your life easier