OASIS High Dimensional Tutorial

This tutorial only assumes familiarity with PRoNTo, but having worked on the classification example of Chapter 13, on the regression example of Chapter 14, and finally on the classification example of Chapter 16 will certainly help your understanding.

In this tutorial we will follow quite closely Chapter 16, where this time we will also add some regression targets to each subject, so that we can also work on a regression problem.

1) Data & Design

In this module you can follow closely the relevant instructions of Chapter 16 with 2 changes.

- In PRoNTo's main window, click on "Data & Design" and a new window will open, "Data and design". Like in previous chapters, browse the directory in which to save the PRT structure (saved as "PRT.mat").
- In the panel Groups, click on "Add and provide a name to the group, e.g. "Demented". We will first add all the information about the demented subjects.
- All the images in the dataset correspond to different subjects; therefore, click on the "Samples" tick box. This will lock the Subjects/Samples field.
- In the "Modalities" panel, click on "Add" and provide a name to the modality, e.g. "MRI_GM", where we used gray matter images.
- In the "Data format" choose the appropriate format for your data (here nifti).
- Load all the image files belonging to the first group. You can select all the files by using the right mouse button and clicking on the option "Select All". Please ensure that the order of the images in the selection is as expected (SPM file selector sorts the files based on their names, which can lead to "Scan10" being selected before "Scan2"). When all the images are selected, click on the "Done" button.
- You can see that the "Design" section is unavailable. Since each subject has only 1 sample/scan, there is no design.
- In the "Regression targets", instead of leaving it as it is, as we did in Chapter 16, this time we will add some values. So click "Specify Targets". After doing so a new window will appear where you can add the list of regression target values. In our case they are available in the "regression_targets_demented.mat" file (*path/OASIS/*) as a .mat file, so you can select it directly by selecting the option "From .mat file" and choosing the appropriate .mat file. After choosing the .mat file you will notice that 2 variables appeared, together with the values for each subject in the "Demented" group. The 2 variables are age and MMSE score. In this tutorial we will try a regression model with the MMSE score. Age here is only for illustration purposes as normally it is either controlled, or regressed out in neuroscience studies.
- Leave the "Covariates" field empty, as we will not focus on regressing out covariates in this tutorial.

After you finish with the "Demented" group, repeat the same procedure with the "NonDemented" group, using the "regression_targets_nondemented.mat" file to add their corresponding regression targets.

- In the "Masks" field, on the bottom left of the "Data and design" window, select the "mask_GM" mask (found inside the OASIS data folder) for the modality specified.
- You can click the "Save" button and then "Quit".

2) Prepare feature set

In this module this time we will not add an atlas, so you just need to provide a name for the feature set, and build the feature set and kernel.

- In PRoNTo's main window, click on "Prepare feature set" and a new window will open, "Prepare feature Set".
- Select the "PRT.mat" file previously created in the "Data & Design" step and another window will open, "Specify modality to include". Select the "Build one kernel per region" tick box, leave all the other default options and click "Done".
- This will bring you back to the "Prepare feature set" window. Provide a name for the feature set.
- Click "Build kernel/data matrix" to build the feature set and kernel.

3) Model: Specify new

In this module we will build one classification model as well as two regression models, one for each group separately. In case there is any confusion on the windows mentioned, feel free to get advice from Chapter 14 from the PRoNTo manual.

For the classification model:

- In PRoNTo's main window, click on "Specify new" and a new window will open.
- Select the "PRT.mat" file and provide a name to the model, e.g. "SVM_GM".
- Select one of the "Feature Set" previously defined. In this case, there is only one: "MRI_GM".
- Leave the option "Use kernels" tick box as it is, i.e. "Yes".
- Select the "Classification" model type and click on the "Define classes" button. A new window will open, "Specify classes", to define the number of classes and a name for each class. We will define 2 classes. For "Class 1" select group "Demented", and all subjects in this first group and, similarly, for "Class 2" select group "NonDemented" and all the subjects in this group. The class names can be any names the user prefers (in alphanumeric characters). Here we simply use the same names as the group names. Leave the "Subsample according to smallest class" as it is for now. This would have been useful if we had unbalanced classes, but in our case we have 50 subjects per group. Once you have appropriately specified everything, click "Done".
- Select the "SVM" option, in the "Machine" field.
- Click the "Optimize hyper-parameter" tick box and leave the range unspecified, i.e. use the default hyperparameter range. Choose "k-folds CV on Subject per Class" for the "Cross-Validation Scheme" (internal loop), and input 5 in the pop up window defining the number of folds.
- Similar to the inner loop, select the "k-folds CV on Subject per Class" cross-validation scheme for the external loop, where k=5.
- In the "Data operations" box, select the "Mean centre features using training data" option.
- Now that we have specified everything about our model you can click on the "Specify and run model" button.

For the two regression models:

- After you click "Specify new" and select the "PRT.mat" file, you can provide a name to the first model, e.g. "KRR_GM_Demented".
- Select from the list one of the "Feature Set" previously defined. In this case, there is only one, "MRI_GM".
- Leave the option "Use kernels" tick box as it is, i.e. "Yes".
- Select the "Regression" model type and click on the "Select subjects/scans" button. This will open a new window, "Specify subjects/scans to regress". For our first model make sure you have chosen all subjects from the "Demented" groups. You can do that by first clicking the "Demented" group and then clicking the "Select all" button to use all subjects from that group. Also choose "mmse" as our regression target (current models available in PRoNTo do no enable multi-output prediction). After you do that click "Done".
- Select the "KRR" option, in the Machine field.
- Click the option "Optimize hyper-parameter" tick box, leave the parameter range with its default values, and choose the "k-folds CV on Subject out" option as a "Cross-Validation Scheme". Once you click the "k-folds CV on Subject out" option a new window will appear where you will be prompted to choose the value of k for the k-folds CV. Type 5 and click OK.
- Do the same for the external cross-validation scheme, as you did for the internal one.
- Finally, in the "Data operations" box, select only the "Mean centre features using training data" option. Click on the "Specify and run model" button, or the "Specify model" if you prefer to only specify all 3 models first and then run them all together.

After you finish specifying all the info about the first model with the subjects from the "Demented" group, you can follow the instructions above but this time choose all the subjects from the "NonDemented" group.

After you've specified all 3 models and you've run them, you are now ready to inspect and study and interpret the results. Feel free to check Chapters 4 and 5 to help you with that.

4) Compute weights

- In PRoNTo's main window, click on "Compute weights" and a new window will open, "Compute weights".
- Select the "PRT.mat" file.
- Select one by one the models you built in the previous step from the list in "Models computed in PRT".
- And finally click on the "Compute weights" button.

Repeat the above procedure for each model you want to compute its weights.

OASIS Low Dimensional Tutorial

This tutorial assumes that you are familiar with PRoNTo and that you have worked on the OASIS High Dimensional Tutorial. Optionally, having worked on the classification example of Chapter 17, particularly the use of .mat files (instead of NIfTI) will certainly help your understanding.

In this tutorial we have used the AAL2 anatomical atlas to transform the high-dimensional brain image of each subject to a low dimensional image. We did that by averaging the values of all voxels of each one of the 116 AAL2 regions, effectively decreasing the dimensions of each subject's image to 116. There are various reasons why an analysis might benefit from lower dimensional data, one of which being the ability to use the primal form of a SVM, instead of the dual one.

1) Data & Design

As with the previous OASIS tutorials, we have sMRI images from 2 different groups of 50 subjects each, the demented group and the non-demented group. So we will start the same way we started before, but now instead of adding the sMRI images of the subjects (which are in NIfTI file format) we will add the summarized images (which are in .mat file format).

- In PRoNTo's main window, click on "Data & Design" and a new window will open, "Data and design". Like in previous chapters, browse the directory in which to save the PRT structure (saved as "PRT.mat").
- In the panel Groups, click on "Add and provide a name to the group, e.g. "Demented". We will first add all the information about the demented subjects.
- All the images in the dataset correspond to different subjects; therefore, click on the "Samples" tick box. This will lock the Subjects/Samples field, please see Chapter 2.4 of the manual for more information on this option.
- In the "Modalities" panel, click on "Add" and provide a name to the modality, e.g. "GM_low_dim".
- In the "Data format" choose the appropriate format for your data, here we have .mat files.
- In the "Files" panel click on the "Select" button, and load all the files belonging to the first group. You can select all the files by using the right mouse button and clicking on the option "Select All", or clicking on the first subject and then Shift-click on the last. Please ensure that the order of the images in the selection is as expected.
- When all the images are selected, click on the "Done" button.
- You can see that the "Design" section is unavailable. Since each subject has only 1 sample/scan, there is no design.
- In the "Regression targets", instead of leaving it as it is, as we did in Chapter 16, this time we will add some values. So click "Specify Targets". After doing so a new window will appear where you can add the list of regression target values. In our case they are available in the "regression_targets_demented.mat" file (*path/OASIS/*) as a .mat file, so you can select it directly by selecting the option "From .mat file" and choosing the appropriate .mat file. After choosing the .mat file you will notice that 2 variables appeared, together with the values for each subject in the "Demented" group. The 2 variables are age and MMSE score. In this tutorial we will try a regression

model with the MMSE score. Age here is only for illustration purposes as normally it is either controlled, or regressed out in neuroscience studies.

- Leave the "Covariates" field empty, as we will not focus on regressing out covariates in this tutorial.

After you finish with the first group, repeat the same procedure for the second group.

- Finally there is the "Masks" field. You can leave the "Masks" field empty, as PRoNTo assumes all .mat files have been preprocessed and masked a priori.
- You can click the "Save" button and then "Quit".

2) Prepare feature set

In this module you just need to provide a name for the feature set, and click "Build kernel/data matrix" to build the feature set and kernel. Follow the instructions of the High Dimensional Tutorial.

3) Model: Specify new

In this module we will build two classification model as well as two regression models. In case there is any confusion on the windows mentioned, feel free to get advice from Chapter 14 from the PRoNTo manual.

For the two classification model:

- In PRoNTo's main window, click on "Specify new" and a new window will open.
- Select the "PRT.mat" file and provide a name to the model, e.g. "SVM_GM".
- Select one of the "Feature Set" previously defined. In this case, there is only one: "GM_low_dim".
- Leave the option "Use kernels" tick box as it is, i.e. "Yes".
- Select the "Classification" model type and click on the "Define classes" button. A new window will open, "Specify classes", to define the number of classes and a name for each class. We will define 2 classes. For "Class 1" select group "Demented", and all subjects in this first group and, similarly, for "Class 2" select group "NonDemented" and all the subjects in this group. The class names can be any names the user prefers (in alphanumeric characters). Here we simply use the same names as the group names. Leave the "Subsample according to smallest class" as it is for now. This would have been useful if we had unbalanced classes, but in our case we have 50 subjects per group. Once you have appropriately specified everything, click "Done".
- Select the "SVM" option, in the "Machine" field.
- Click the "Optimize hyper-parameter" tick box and leave the range unspecified, i.e. use the default hyperparameter range. Choose "k-folds CV on Subject per Class" for the "Cross-Validation Scheme" (internal loop), and input 5 in the pop up window defining the number of folds.
- Similar to the inner loop, select the "k-folds CV on Subject per Class" cross-validation scheme for the external loop, where k=5.
- In the "Data operations" box, select the "Mean centre features using training data" option.
- Now that we have specified everything about our model you can click on the "Specify and run model" button.

After you finish specifying all the info about the first classification model with the kernel SVM machine, you can follow the instructions above but this time uncheck the "Use kernels" tick box so that we now have the machine operating in the primal form. And now instead of choosing SVM you can instead choose Binary L2-SVM. The rest of the options are the same as above.

For the two regression models:

- After you click "Specify new" and select the "PRT.mat" file, you can provide a name to the first model, e.g. "KRR_GM_Demented".
- Select from the list one of the "Feature Set" previously defined. In this case, there is only one, "MRI_GM".
- Leave the option "Use kernels" tick box as it is, i.e. "Yes".
- Select the "Regression" model type and click on the "Select subjects/scans" button. This will open a new window, "Specify subjects/scans to regress". For our first model make sure you have chosen all subjects from the "Demented" groups. You can do that by first clicking the "Demented" group and then clicking the "Select all" button to use all subjects from that group. Also choose "mmse" as our regression target (current models available in PRoNTo do no enable multi-output prediction). After you do that click "Done".
- Select the "KRR" option, in the Machine field.
- Click the option "Optimize hyper-parameter" tick box, leave the parameter range with its default values, and choose the "k-folds CV on Subject out" option as a "Cross-Validation Scheme". Once you click the "k-folds CV on Subject out" option a new window will appear where you will be prompted to choose the value of k for the k-folds CV. Type 5 and click OK.
- Do the same for the external cross-validation scheme, as you did for the internal one.
- Finally, in the "Data operations" box, select only the "Mean centre features using training data" option. Click on the "Specify and run model" button, or the "Specify model" if you prefer to only specify all 3 models first and then run them all together.

After you finish specifying all the info about the first model with the subjects from the "Demented" group, you can follow the instructions above but this time choose all the subjects from the "NonDemented" group.

After you've specified all four models and you've run them, you are now ready to inspect and study and interpret the results. Feel free to check Chapters 4 and 5 to help you with that.

4) Compute weights

- In PRoNTo's main window, click on "Compute weights" and a new window will open, "Compute weights".
- Select the "PRT.mat" file.
- Select one by one the models you built in the previous step from the list in "Models computed in PRT".
- And finally click on the "Compute weights" button.

Repeat the above procedure for each model you want to compute its weights.