CO416 – Machine Learning for Imaging

Coursework 2

Getting started

You are given a main Jupyter notebook that contains descriptions and some skeleton code. You will need to add your implementations at the specified locations directly within the notebook.

More details about this coursework can be found on page 2 of this spec sheet.

Step 1: Set up your virtual machine (see tutorial 5)

If you haven't done already, you need to set up your GPU virtual machine following the instructions given in tutorial 5. This includes cloning of the coursework 2 repository to your VM. The repo is here: `https://gitlab.doc.ic.ac.uk/bglocker/mli-coursework-2`

Step 2: Learn basics of medical image computing

In the coursework repository, you will find a notebook named **`MLI-MIC-Summary.ipynb`**. This notebook is an introduction to medical image computing with many functions and algorithms that will be useful for the main coursework tasks. Go through the notebook from top to bottom. You will not need to implement anything in this notebook, all model answers are provided. But, of course, feel free to play around and change any settings or parameters to study their effect.

Step 3: Work on the main coursework tasks

The main notebook for the coursework is **`MLI-CW-2.ipynb`**. Read carefully through the instructions given in the text cells. The notebook should contain all information you need to finish each task. You are asked to finish all tasks which are highlighted in the notebook.

Start early, and ask questions if anything is unclear!

SUBMISSION

You will need to upload an **archive** (`.zip` or `.tar.gz`) with the base name `coursework2` to CATE. The archive should contain **both the notebook .ipynb as well as a .pdf exported version of it**. You can export a PDF version of your notebook from Jupyter via the 'Download as…' menu.

Submission deadline on CATE: Thursday, 28 February, 23:59 (midnight)

Age regression from brain MRI

Predicting age from a brain MRI scan can have diagnostic value for a number of diseases that cause structural changes and damage to the brain. Discrepancy between the predicted, biological age and the real, chronological age of a patient might indicate the presence of disease and abnormal changes to the brain. For this we need an accurate predictor of brain age which may be learned from a set of healthy reference subjects.
The objective for the coursework is to implement two different supervised learning approaches for age regression from brain MRI. Data from 600 healthy subjects will be provided. Each approach will require a processing pipeline with different components that you will need to implement using methods that were discussed in the lectures and tutorials. There are dedicated sections in the Jupyter notebook for each approach which contain some detailed instructions, hints and notes.

Part A: Volume-based regression using brain structure segmentation

The first approach aims to regress the age of a subject using the volumes of brain tissues as features. The structures include grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF). It is known that with increasing age the ventricles enlarge (filled with CSF), while it is assumed that grey and white matter volume may decrease over time. However, as overall brain volume varies across individuals, taking the absolute volumes of tissues might not be predictive. Instead, relative volumes need to be computed as the ratios between each tissue volume and overall brain volume. To this end, a four-class (GM, WM, CSF, and background) brain segmentation needs to be implemented and applied to the 600 brain scans. Brain masks are provided which have been generated with a state-of-the-art neuroimaging brain extraction tool. Different regression techniques should be explored, and it might be beneficial to investigate what the best set of features is for this task. Are all volume features equally useful, or is it even better to combine some of them and create new features. How does a simple linear regression perform compared to a model with higher order polynomials? Do you need regularisation? How about other regression methods such as regression trees or neural networks? The accuracy of different methods should be evaluated using two-fold cross-validation, and average age prediction accuracy should be compared and reported appropriately.

Part B: Image-based regression using grey matter maps

The second approach will make use of grey matter maps that have been already extracted from the MRI scans and aligned to a common reference space to obtain spatially normalised maps. For this, we have used an advanced, state-of-the-art neuroimaging toolkit, called SPM12. The reference space corresponds to the commonly used MNI atlas as seen in the lecture on image segmentation. Because these grey matter maps are spatially normalised (ie., registered), voxel locations across images from different subjects roughly correspond to the same anatomical locations. This means that each voxel location in the grey matter maps can be treated as an individual feature. Because those maps are quite large at their full resolution there would be a very large number of features to deal with (more than 850,000). A dimensionality reduction using PCA may need to be performed before training a suitable regressor on the low-dimensional feature representation obtained with PCA. It might also be beneficial to apply some pre-processing (downsampling, smoothing, etc.) before running PCA, which should be explored. The implemented pipeline should be evaluated using two-fold cross-validation using the same data splits as in part A, so the two different approaches can be directly compared in terms average age prediction accuracy.