Initial Plan: Machine learning: Predicting Age and Cognitive Performance from MRIs

by Craig Winfield

Supervisor: Matthias Treder

Module Number: CM3203
Module Title: One Semester Individual Project
Credits for Module: 40

February 3rd 2019

Ethics

The data set for this project has been kindly provided by Cam-CAN. Cam-CAN is the Cambridge Centre for Ageing and Neuroscience, a large-scale collaborative research project. The Cam-CAN project is using epidemiological, behavioural, and neuroimaging data to understand how individuals can best retain cognitive abilities into old age. The Cam-CAN data set is within the public domain, and an access form has been signed and sent off to give permitted access to the data set for the intents and purposes of this project. The data set will be analysed through a password and security protected PC at Cardiff University with GPU-enabled tensorflow.

The Cam-CAN data set includes the following types of data:

- Brain structure Measures
- General Health Measures
- Demographics data
- Cognitive Performance

Project Description

The average age of society is steadily increasing, therefore we need to look for earlier signs of deterioration in cognitive ability. The earlier that abnormalities are spotted, the sooner and better they can be treated, and MRIs can spot early signs of disease before any symptoms become visible. From a scientific point of view, this project provides an exciting opportunity to make steps towards capturing the relationships between brain tissue and cognitive variables, and provide further insight into cognitive function with regards to age. In this project, I aim to characterize age and behaviourally-relevant features using structural MRIs (Magnetic Resonance Imaging) as inputs. An MRI is a 3D medical imaging technique used in radiology to form pictures of the anatomy and the physiological processes of the body in both health and disease. MRI scanners use strong magnetic fields, magnetic field gradients, and radio waves to generate images of the organs in the body. Examples of physical changes include gray-matter loss, the brain shrinking, and the decline of white-matter.

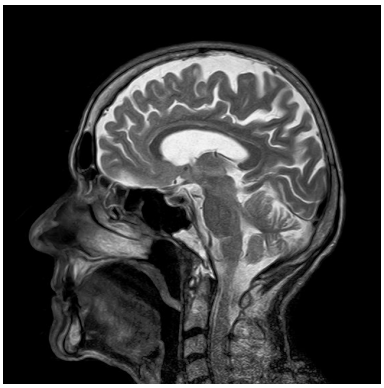

Figure 1: E.g. MRI

I will be developing a Predictive model of Machine Learning. Predictive modeling includes a Machine Learning algorithm to learn properties from a training data set to make predictions. Predictive modelling can be divided further into the area of Regression, where a Regression model is based upon the analysis of relationships between variables and trends to make predictions about variables. In the context of the project, the model will be able to predict age and cognitive performance by detecting signs of ageing in an MRI. The model will have to be evaluated to measure how successful it is at predicting variables from given data. Cross-Validation is a method of evaluating the effectiveness of a model, given a data set. Cross-Validation partitions the data set into subsets, then performs analysis on one subset. After that, analysis is done on other subsets. Data is split into two sets: training and test. The training set will train

the model, while the test set is used to measure how well the model performs at making predictions on that test set (which acts as unseen data). To reduce variability of results, numerous rounds of cross-validation are carried out, using different partitions of the data set. An average of the results is then generated and cross-validation combines (averages) measures of fitness in prediction to derive a more accurate estimate of model prediction performance. By applying this technique to different models, we can see which predictive models are more accurate, based on their combined measures of fitness.

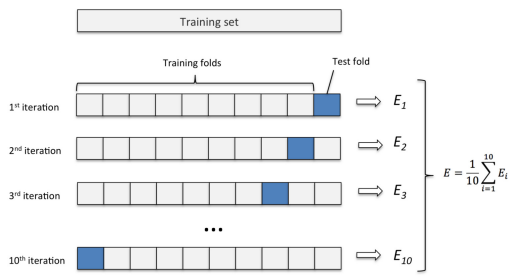

Figure 2: Cross-Validation with 10 iterations and 10 partitions of data

Inspired by 2D and 3D convolutions, the aim of this project is to use machine learning on the Cam-CAN ageing neuroscience data-set. Due to the non-linearity nature of ageing, it's hard to comprehensively model using standard linear statistical techniques such as GLM (Generalized Linear Models). The goal will be to predict the age of participants and physiological variables from MRIs. Augmentation of linear models by e.g. quadratic terms only partially address this problem since it fails to address other relationships. As an alternative, deep neural networks are known to learn arbitrary non-linear relationships. In recent years, Convolutional Neural Networks (CNNs) have been very successful at extracting features from raw images thereby circumventing the necessity for prior feature selection / dimensionality reduction through methods such as ICA (Independent Component Analysis).

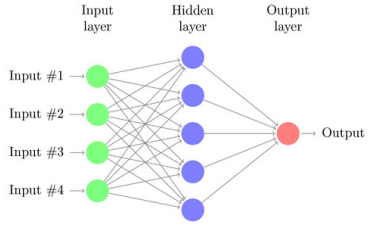

Figure 3: Neural Network with one hidden layer

Project Aims & Objectives

The aims of the project are based around developing a Convolutional Neural Network to predict age and cognitive dissonance, given raw MRIs as input. This predictive model will detect changes in structural MRIs related to ageing, where changes include the loss of gray-matter, white-matter, and volume in the brain, common signs of ageing of the brain. The aims are outlined as follows:

Aims

- Develop a CNN to closely predict the age of a given MRI, trained via the Cam-CAN data set.
- Develop a CNN to closely predict the cognitive dissonance of a given MRI, trained via the Cam-CAN data set.

It is worth noting that my first priority will be to develop the system to predict Age, and if time allows it, I shall further develop the system to predict Cognitive Dissonance. To tackle this project, I will be using different techniques and ideas used in standard Data Science-esque tasks. I have listed the following objectives that suggest how I will be able to achieve the above aims of this project:

Objectives

- Research regression models and evaluate which model is best for the problem. Such models include linear regression and support vector regression
  - Identify which regression models are appropriate for the problem of this project.
  - Evaluate the effectiveness of models using T Tests to make inferences about data beyond our current set of data, i.e. making inferences with regards to ageing beyond just the sample I am testing.
  - Evaluate the effectiveness of models using ANOVA to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups e.g. Splitting participants from data set into groups based on a selected criteria. Could suggest a massive difference in MRI as a result of ageing between older and younger population.
- Reduce Dimensionality of MRI images from 3D to 2D to feed as data input for Convolutional Neural Network.
- Develop a Convolutional Neural Network (CNN) to to extract features from a raw image at 2D and 3D Dimensionality

- Develop a CNN that with a high test set success rate from appropriate testing techniques.

Work Plan

Deliverables

I endeavour to submit the following items by the deadline:

- Final report
- All source code
- All supporting documents related to my findings and analysis
- Any data visualisation related to the project.

Research and Preparation

Before getting started with development, I need to read up materials relevant to the project. I will explore tutorials to understand how to properly use the frameworks that will be used for this project, e.g. Tensorflow. Once I have a feel for the frameworks, I can start to make early strides in developing a solution. Furthermore, I need to understand which Linear Regression Model is most suitable for this type of problem. Different models will give differing levels of performance so it is important I evaluate the performance of the models. For evaluation, I will use testing techniques such as Cross-Validation to measure a model's performance. A model's performance will be defined by the success rate of training data to test data.

Implementation

After spending time researching and testing ideas, I'll start implementing ideas I have come across through my time researching. Such materials can be related to MRIs or Machine Learning programming. I'm going to be implementing a Convolutional Neural Network, which I will experiment with different linear regression models to see which provides the best performance for the project and data set. Initially to get started, the MRIs have to be sliced from 3D to 2D images to get started in training and developing a neural network.

Supervisor Meetings

I have scheduled numerous weekly meetings with my supervisor, Matthias Treder, but I won't have a meeting every week. These meetings are an opportunity to share progress, raise issues, seek help, and come together with other students who are working on this project. There will come a time where Matthias Treder will be away, so meetings will be conducted through an appropriate medium like Skype.

Milestones

Taking into consideration commitments to my other University module, I have aimed to reach these weekly milestones by the end of the labelled week :

- **Week 1** 28/01/19-03/02/19:
  - Download all necessary software on my personal laptop. List of software is not limited to but includes: Anaconda, Tensorflow, Keras, MRIcroGL.
  - Initial Meeting with Supervisor and Project Team to discuss next steps to make with the project. Workshop with Supervisor to obtain the Cam-CAN data and start experimenting with the data. Using Anaconda to take slices of MRIs that can be put into a Convolutional Neural Network. This reduces the dimension from 3D to 2D in order to get started.
  - Writing up initial plan
- **Week 2** 04/02/19-10/02/19:
  - Submit Initial Plan of the project. Initial Plan would have been discussed from previous one to one meeting with supervisor to properly understand what to write in the initial plan.
  - Implement a Neural Network via tutorials to show understanding towards development
- **Week 3** 11/02/19-17/02/19:
  - Program a 2D Convolutional Neural Network to predict age, and get training and test error. Produced through the use of data science frameworks via Anaconda. Develop either with slices of Cam-CAN data, or by using toy data
- **Week 4** 18/02/19-24/02/19:
  - Program a 2D Convolutional Neural Network to predict age, and get training and test error. Produced through the use of data science frameworks via Anaconda. Develop either with slices of Cam-CAN data, or by using toy data
- **Week 5** 25/02/19-03/03/19:
  - Program a 3D Convolutional Neural Network to predict age, and get training and test error. Produced through the use of data science frameworks via Anaconda. Develop either with slices from Cam-CAN data, or by using toy data
- **Week 7-8** 11/03/19-24/03/19:

  - Start implementing solution to the project, based on research and design carried out in prior weeks. Focus on tackling Age prediction first.
  - Measuring and analysing the performance of models using appropriate testing techniques, such as Cross-Validation

- **Week 9** 25/03/19-31/03/19:
  - Experiment with different linear regression models and analysis the best fit for the project.
  - Aim to narrow down to at most best 2 choices of regression models, if not a sole best choice.

- **Week 10-11** 01/04/19-14/04/19:
  - Start evaluating the predictive modelling through the use of Cross-Validation to measure the strength of the models predictions
  - Produce a Convolutional Neural Network that can predict Age from MRI data input. Produced through the use of data science frameworks via Anaconda. Developed to work with the Cam-CAN data, the Neural Network should take in unseen MRIs to correctly predict age.

- **Week 12-13** 15/04/19-28/04/19:
  - Produce a Convolutional Neural Network that can predict Cognitive Dissonance from MRI data input. Developed to work with the Cam-CAN data, the Neural Network should take in unseen MRIs to correctly predict Cognitive Dissonance.

- **Week 14** 29/04/19-05/05/19:
  - Write up report on progress so far, including necessary results, screenshots, and analysis.

- **Week 15** 06/05/19-10/05/19:
  - Finalizing solution for the project. Testing, debugging, refactoring code.
  - Hand in final report.