Age Regression from Brain MRI
Group: 22

Jeremy Xie, Tianquan Zhang, Haotian Zhu
{linge.xie19, tianquan.zhang19, haotian.zhu19}@imperial.ac.uk

1 Summary

In this project, we were tasked to investigate the possibility of implementing different supervised learning approaches for age regression from brain MRI with a total of 652 healthy samples for training and 100 subjects for testing. In Part A, we were given the assignment of first segmenting the brain MRI images followed by applying multiple regression methods to predict the ages of the subjects. In the second part, we applied the similar regression methods after pre-processing and PCA dimensionality reduction. In the last part, we explored CNN-based architectures for better regression performance.

2 Part A

In this part, we were given the task of performing several regression methods that would help us to predict the ages of the subjects after image segmentation techniques were performed. Here for the initial portion, we pre-processed the images using the functions provided in the documentation by first converting the mean to 0 and variance to 1. Then we resampled the images to a new size and new spacing of $[1., 1., 1.]$

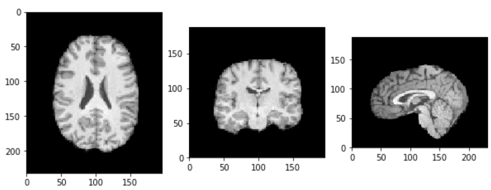

Figure 1: MRI image before processing

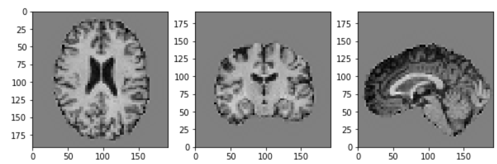

Figure 2: MRI image after processing

For the segmentation, we proposed a simplified version of U-Net that performed downsampling twice using 3-dimensional Max Pooling layers and upsampling twice using 3-dimensional Transpose Convolutional layers with two 3-dimensional Convolutional layers as bridge. We have reached a cross-entropy loss of 0.288151 after 100 epochs. Here we have acquired a test loss of 0.300243, **all the results with Dice scores could be found in part 5**. Following the segmentation, we calculate three absolute brain tissue volumes for each subject against ages with normalisation. The boxplot for Dice scores and the scatter plots for tissue volumes are displayed as follows. The shape for X and y are $(500, 3)$ and $(500, 1)$ respectively

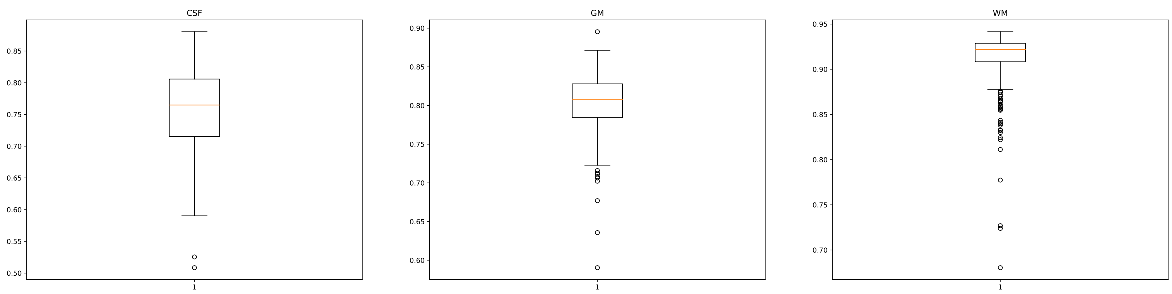

Figure 3: Box-and-whisker plot for Dice Similarity Coefficient for class CSF,GM and WM from left to right

The remaining section of part A demanded 3 different regression approaches for age prediction using the input we have acquired, here we have selected to perform Linear Regression, Bayesian Ridge Regression and Decision Tree

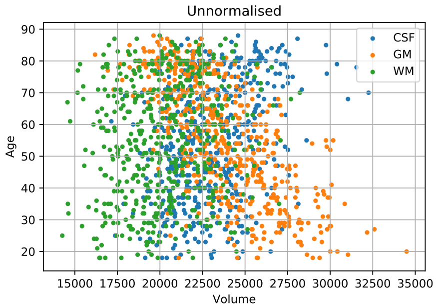

Figure 4: Unnormalised scatter plots for tissue volumes

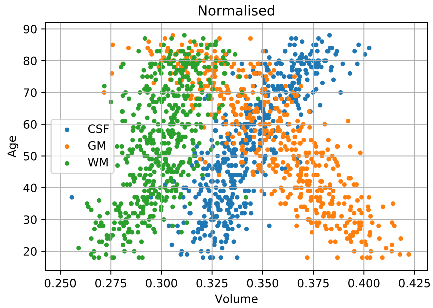

Figure 5: Normalised scatter plots for tissue volumes

Regression from the `scikit-learn toolkit`. We also use the `KFold` function to perform a 2-fold cross validation that splits the dataset of 500 into equally sized training and testing sets. Each training and testing sets were used exactly once.

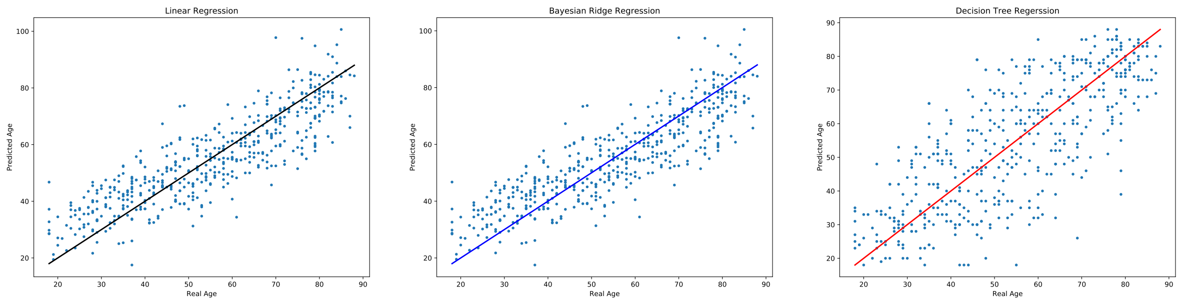

Figure 6: Scatter plots for 3 different regression methods in Part A with LR, BRR and DR from left to right

3 Part B

In order to pre-process the spatially normalised GM maps, we applied Discrete Gaussian first with sigma value of 0.8 followed by an image-based normalisation. The final procedure involves using Recursive Gaussian image filter provided inside the `sitk` toolkit with factor and sigma calculated using scale of 2.

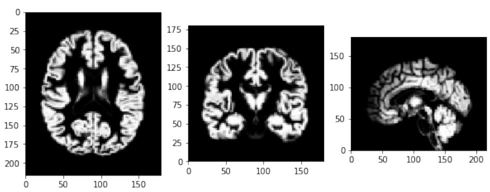

Figure 7: GM maps before pre-processing

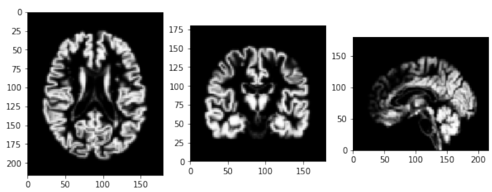

Figure 8: GM maps before pre-processing

Here we have adopted the PCA in the second portion from the `scitkit-learn toolkit`. We have set the *n_components* to 0.95 for our pca model as to preserve 95% of the variance and fit it to the training data using the fit function. Then we performed dimensionality reduction techniques using the transform function to both training and test data. The remaining component is similar to A-3 with the last regression selection for us changed to Logistic Regression. The new shape of X_{train} and X_{test} are (250, 874800) and (250, 227) respectively. The plots for real age vs. predicted age can be found at 9

4 Part C

In this section, we load the spatially normalised GM maps and upsampled the images from $[64, 64, 64]$ to $[96, 96, 96]$. We have adopted the architecture of `Le-Net` such that it involves 3 convolutions followed by 3 Linear layers that eventually output a regression result instead of an image. During training, we perform a 2-fold cross validation with loss metrics changed to l1_loss. L1_loss calculate the MAE automatically that best fits our needs in this task. We proposed a few models, however, none has out-performed the modified LeNet network. The results could be seen in part 5.

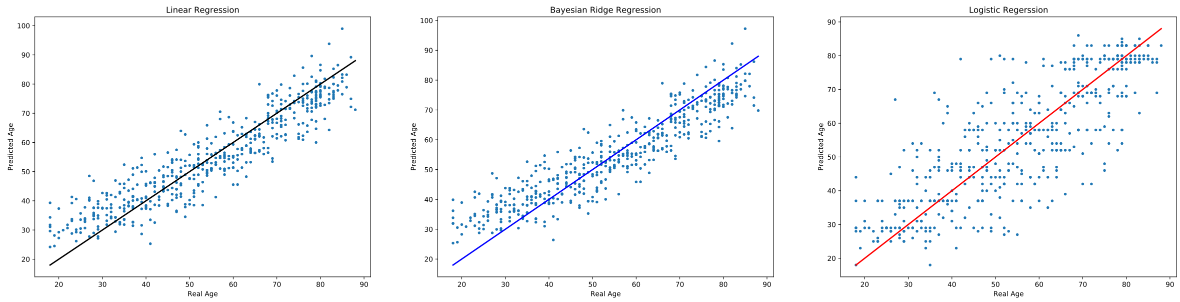

Figure 9: Scatter plots for 3 different regression methods in Part B with LinearR, BRR and LogisticR from left to right

5 Age Regression Results

The metrics we used are Mean Absolute Error and R2 score, here we have retrieved our results with hyper-parameters of *batch_size* of 4, *learning_rate* of 0.001, *num_classes* of 4 for part A and B. We used the same *batch_size* for part C. All the results were acquired using 100 epochs.

	MAE	R2 Score
Linear Regression	7.5754375	0.7327915
Bayesian Ridge Regression	7.5188744	0.7375202
Decision Tree Regression	9.892	0.5128438

Table 1: Part A metrics(training)

	MAE	R2 Score
Linear Regression	5.7808416	0.8463146
Bayesian Ridge Regression	6.0743911	0.8340973
Logistic Regression	8.402	0.6430588

Table 2: Part B metrics(training)

	Dice Similarity Coefficient
CSF	0.7578
GM	0.8048
WM	0.9144

Table 3: Part A Dice scores(training)

	MAE
LeNet	5.686756

Table 4: Part C metrics(training)

As we can see the best regression methods for part A is Bayesian Ridge Regression and Linear Regression for part B. For the last task, after two-fold cross validation with 100 epochs each, we have reached a MAE of 5.686756 at validation.

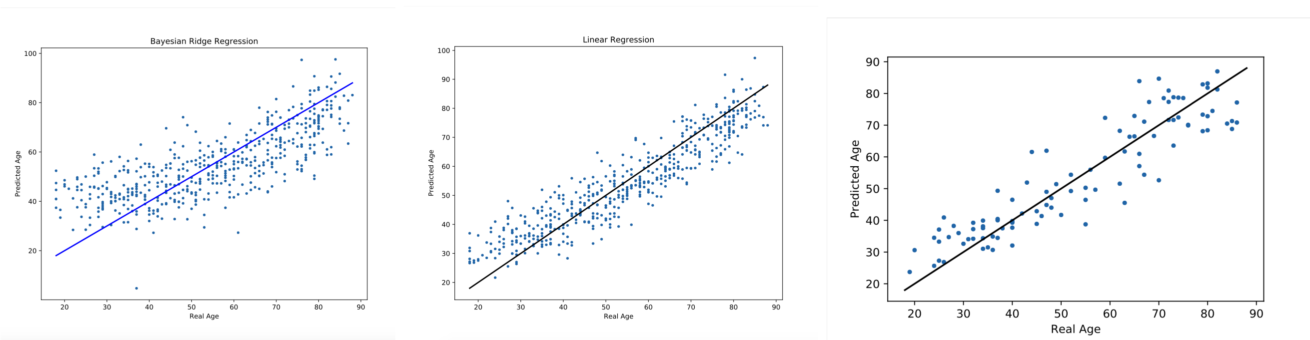

Figure 10: Scatter plots for age regression for three different approaches A, B and C from left to right

At the last step, we loaded the actual test set of 100 subjects that were released on February the 24th and we have the following metrics re-calculated for the test set.

	MAE	R2 Score
Linear Regression	10.6583	0.5709
Bayesian Ridge Regression	10.6640	0.5704
Decision Tree Regression	13.25	0.2857

Table 5: Part A metrics(test)

	MAE	R2 Score
Linear Regression	5.9135	0.8630
Bayesian Ridge Regression	6.1988	0.8481
Logistic Regression	8.29	0.7022

Table 6: Part B metrics(test)

	Dice Similarity Coefficient
CSF	0.7492
GM	0.8011
WM	0.9122

Table 7: Part A Dice scores(test)

	MAE	R2 Score
LeNet	6.106610	0.844344

Table 8: Part C metrics(test)

It is clear that our model performed rather poorly when it compares to the training set with the test results of MAE from all methods raised significantly in part A. This gives us some possibilities to further develop our model by increasing its capacity as we have reached a test set loss of **0.323529**. The amount of increase in terms of MAE in part B is rather trivial comparing to part A, this is largely due to the success of dimensionality reduction, performed by our PCA model. We have seen the MAE raised in the last section, this is normal due to the fact that test error is the generalization error of the model. If the training error is incredibly smaller than the test error, it could be due to the variance of the data or the model is overfitting, however, this is not our case here.