Gender-Based Age Estimation using Quantified Brain Asymmetry

Bodhisatwa Chatterjee

June 5, 2019

Abstract

In this term project, we have revisited the idea of age-estimation using quantified brain asymmetry with the perspective of gender. We have included the factor the gender of the subject while designing regression models for human-age estimation. Broadly, we are interested into two questions: whether gender of a subject influences the estimation of age and whether male and female have different asymmetry. After performing the classification task and then training the regression model has given us superior insights to both the questions. We have uncovered about 200 distinct features which can distinguish the gender of human subjects and employing separate regression models resulted in improved accuracy of age estimation in various splits.

Contents

1 Introduction

In this term project, we have revisited the problem of estimation of human age using quantified brain asymmetry in MR Images. Specifically, we have combined the existing model of age estimation [1] with a classification model which will first learn to categorize the samples as *male* or *female* and then we have developed separate regression models for the male and female subjects for the age estimation, after doing feature selection for both the problems.

The reason for following such an approach is that we believe that for accurate estimation of human ages from quantified asymmetry of brain in MR images, we need to account the factor of gender in the subjects. The core of this belief arises from the fact that male and female brain have structural differences[2][3], which might lead to different asymmetry of the brain in both the genders. This might result in having different features for learning the regression model for age estimation. Therefore, we should train different regression models for age estimation of male and female subjects.

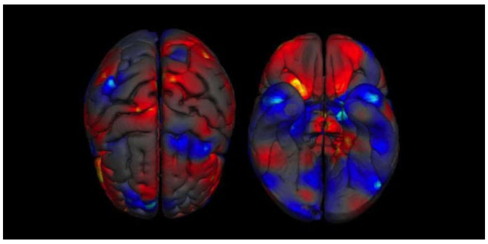

Figure 1: Anatomical Differences Between Male and Female Brain[3]

Problem Statement Train a binary classification model which would categorize subjects into two categories as *male* or *female*. Based on the separate categories, train a regression model based on quantified asymmetry of human brain for each category for estimation of age. Compare the results of these regression models with the result of the original regression model(gender-neutral) model.

2 Approach

In this section, we would describe the dataset on which we are working on. We would also present the specific tasks we are looking at, namely classification and regression. A new technique called *Reverse Feature Selection* has also been described in this section, which precedes the classification task.

2.1 Data

The dataset which we are using is the *Normal Aging Dataset*. This dataset has information for 198 subjects: 100 females and 98 males. There are three class labels in the original dataset: 1,2 and 3, each of which corresponds to whether the subject's age is in fifties,sixties or seventies respectively. The original dataset had 55 million features. A summary of this dataset has been presented in Table 1. In this version of the dataset, we

Total Subjects	Males	Females	Total Features	Features after Selection	Classes
198	98	100	55,000,000	1000	3

Table 1: Summary of the Normal Aging Dataset

have 1000 features which are ranked according to their augmented variance ratio(AVR). The split has already been provided inside the dataset: *Leave-12-out* cross validation for males and *Leave-13-out* cross validation for females. To make sure that this dataset is suitable for the classification task, we have added an extra label of columns in the data. The label 0 denotes the males and label 1 denotes the females. This will be useful for the classification task of determining genders.

2.2 Methods

In this section, we will describe the exact procedures which were followed for the this project.

2.2.1 Creation of Modified Dataset

The dataset provided to us was did not have any labels for determining genders. We first took the separate splits in each folder of male and female portions of the dataset and added one column of extra labels to act as the dataset for the binary classification problem of determining gender. The matlab modules *Load_Data_F.m* and *Load_Data_M.m* created in the separate 'male' and 'female' folder.

Once separate labels have been added, the separate splits for male and female portion of the dataset have been combined to create a unified dataset. This dataset contains the mixture of male and female subject, separated by 0/1 binary class labels. The matlab module *Gen_Data_Set.m* combines these datasets and its output has been copied to the directory named *Combined_Dataset*. All the machine learning tasks used in this project operates on this dataset.

2.2.2 Reverse Feature Selection

The original 1000 features plus the original class labels(1,2,3) results in 1001 features in total for this dataset. However, these 1001 features were originally designed for the human-age estimation problem via regression model, not for the classification task which establishes the gender of the subject. Therefore, if we train on the complete 1001 features for the classification task, we would get abnormal value of accuracy and standard deviation across all the classification tasks. Therefore, we would need to run by classification task by taking each features one by one and removing those features which are not suitable for this binary classification task. This is called **Reverse Feature Selection**, where we remove the features which gives us abnormal accuracy. This algorithm is similar to the backward selection algorithm, where we have the whole feature set and iteratively remove features one by one. The matlab module *Check_Features.m* and *forwardselsection.m* performs this task.

2.2.3 Building the initial classifer

Owing to the comparatively straight forward binary classification problem of gender categorization, we have used decision tree for the actual classification. The classification accuracy, classification matrix and confusion matrix have been shown in the results section. The matlab module *'Build_Classifier.m'* is for this task.

2.2.4 Reconstructing the Dataset

From the results of the decision tree classifier, we now have categories of male and female in both the original train and test split. The features which were removed for the classification task has to brought back in the dataset for fitting of regression models. A backup copy of the original feature matrix was stored before starting the classification task and is restored in this step.

2.2.5 Training the Regression Models

Based on the results of the Decision Tree Classifier, we have two categories: Males and Females. Two separate linear regression models were trained on the output of the classifier. The matlab module *'fitRegressionModel.m'* contains the implementation of fitting the linear regression models.

3 Results

In this section, we will present the results for classification and regression. We will also provide the original linear regression model which would provide us a baseline to compare our results

3.1 Classification Results

The Train and Test Accuracy of the binary classification problem of categorizing subjects into their respective genders. These results were achieved over the default 51 splits provided in the dataset.

Train Accuracy	Train Stddev	Test Accuracy	Test Stddev
0.9960	0.067	0.9543	0.032

Table 2: Test and Train Accuracy for categorizing gender of test subjects

Test Classification Matrix	
0.9950	0.0050
0.0031	0.9969

Table 3: Average Classification Matrix for categorizing gender

Test Confusion Matrix	
11.9400	0.0600
0.0400	12.9600

Table 4: Average Confusion Matrix for categorizing gender

3.2 Regression Results

In this section, the plots of training separate regression models for male and female subjects to estimate age has been provided. Every male/female model is accompanied by the original regression model, which is the unified model for estimating age, regardless of gender. The testing was done for all the 51 splits, the plots are showed for splits 14-18 in Fig 2 to Fig 16.

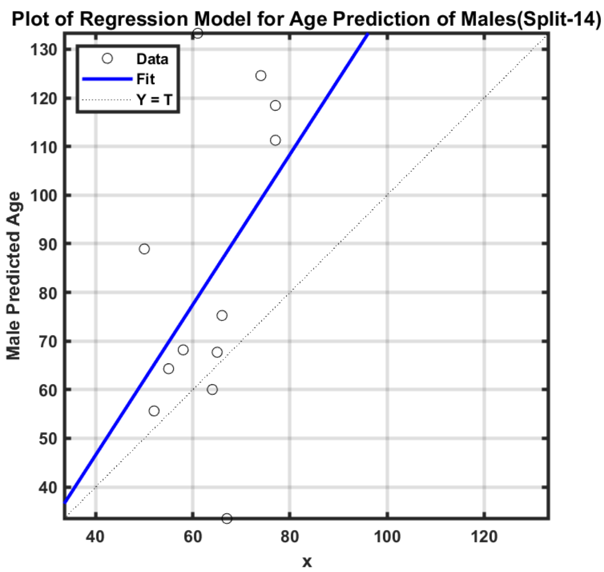

Figure 2: Regression Model for Male Age Prediction(Split-14)

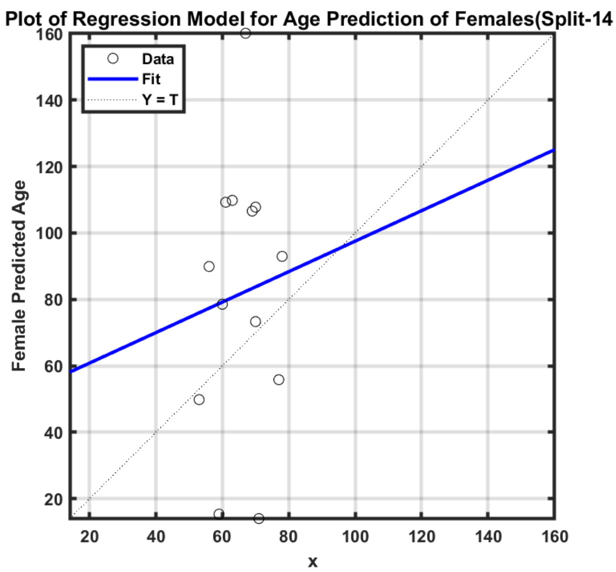

Figure 3: Regression Model for Female Age Prediction(Split-14)

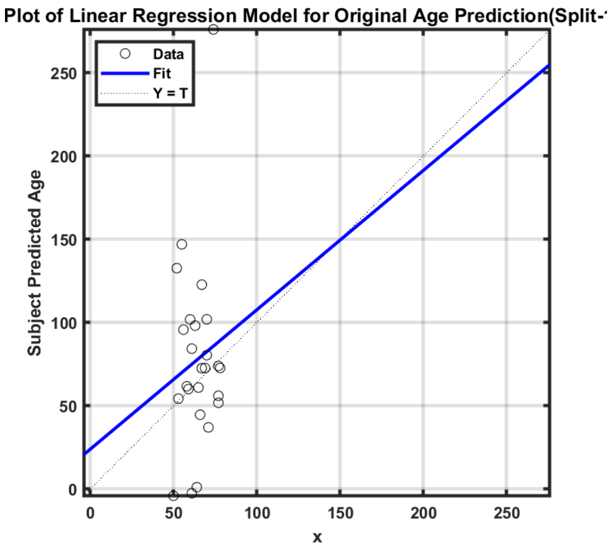

Figure 4: Original Gender Neutral Regression Model(Split-14)

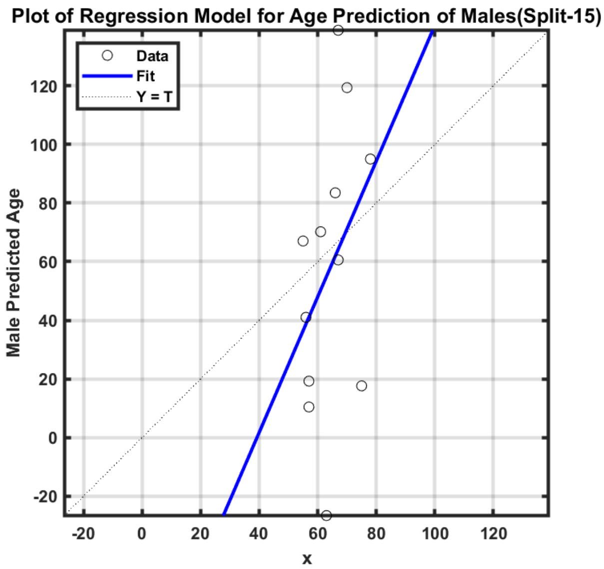

Figure 5: Regression Model for Male Age Prediction(Split-15)

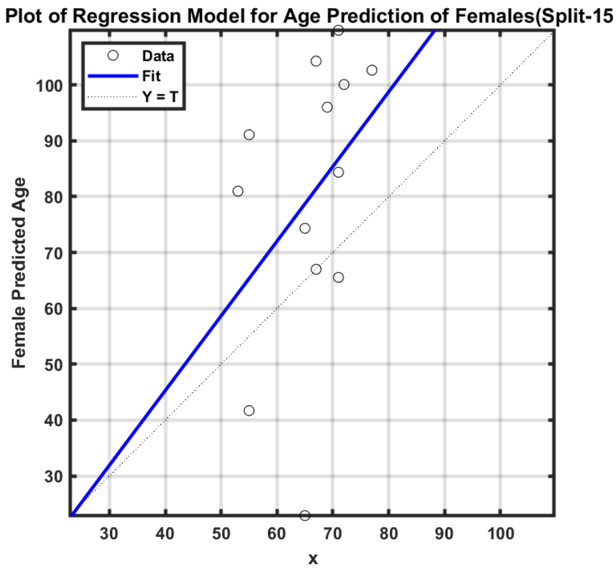

Figure 6: Regression Model for Female Age Prediction(Split-15)

4 Related Work(State of Art)

Estimation of human age from MR images of brain is a very common approach. Most of the approaches have to goal to identify anomalies in the structure of brain for the early detection of neurodegenerative diseases like Alzheimer's disease (AD) or Parkinson's disease(PD).

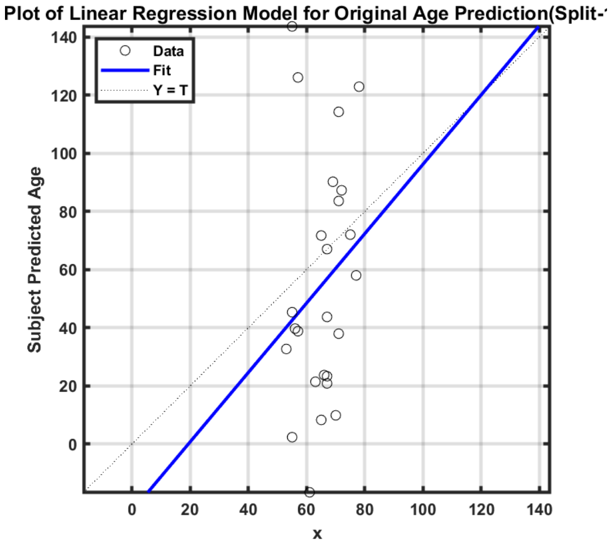

Figure 7: Original Gender Neutral Regression Model(Split-15)

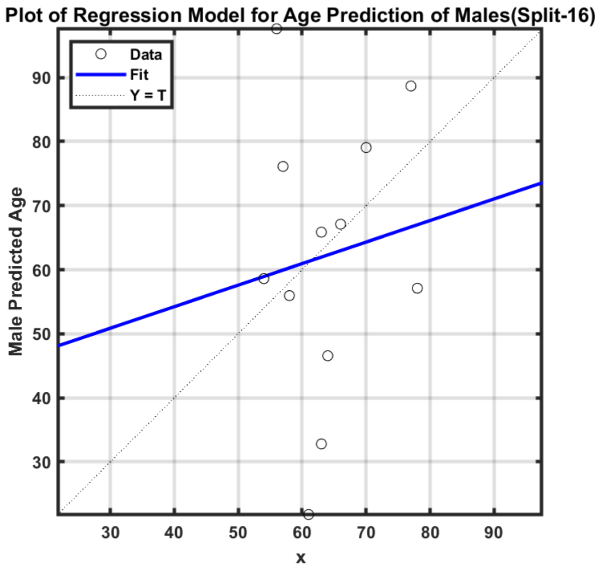

Figure 8: Regression Model for Male Age Prediction(Split-16)

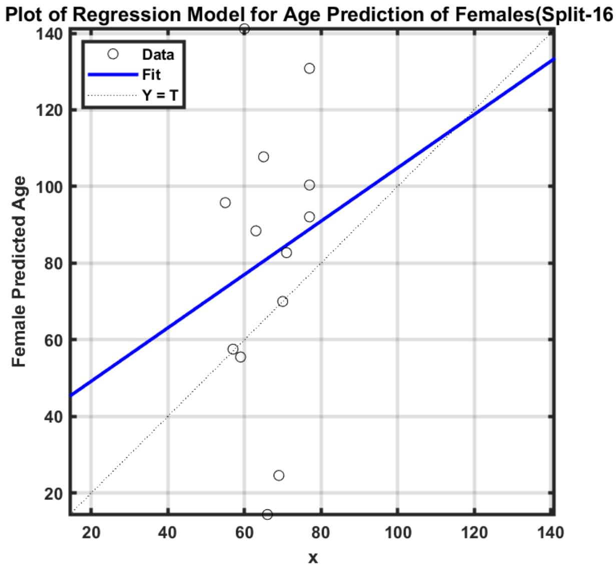

Figure 9: Regression Model for Female Age Prediction(Split-16)

4.1 Related Work on the dataset area

There are many related work using MR images of brain; some of which has the objective of age estimation and most of them don't. Popular works using MR images of brain are [4], [5] which deals with segmentation of MR images of brain. [5] uses Markov Random Fields to segment images of T1-weighted brain MR images, while [4] also employs Markov Random Field for image segmentation for age estimation, the preprocessing techniques are entirely different.

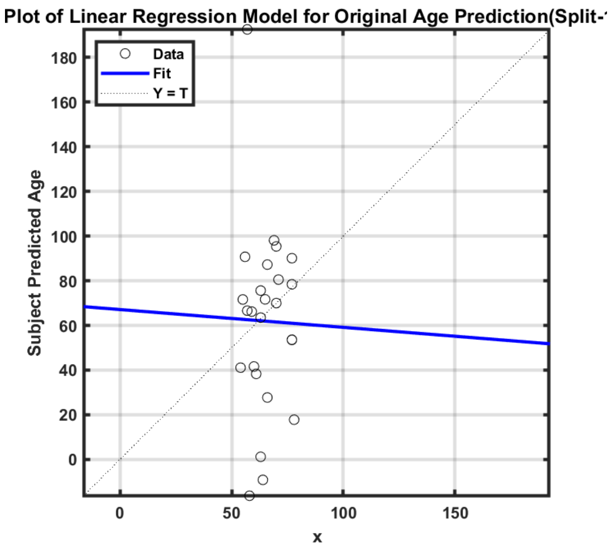

Figure 10: Original Gender Neutral Regression Model(Split-16)

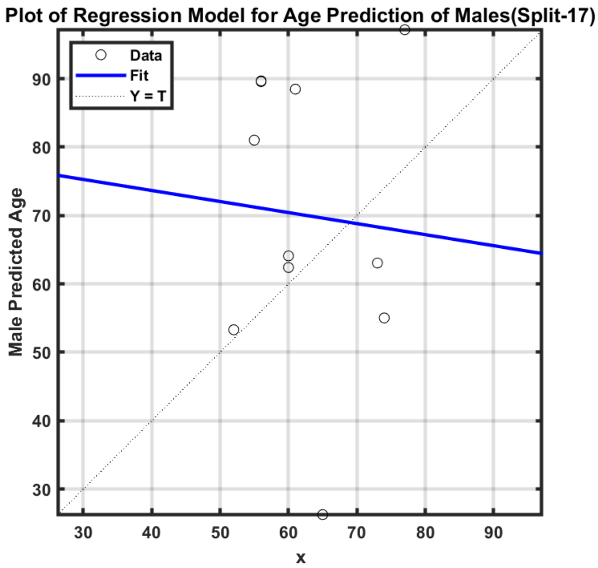

Figure 11: Regression Model for Male Age Prediction(Split-17)

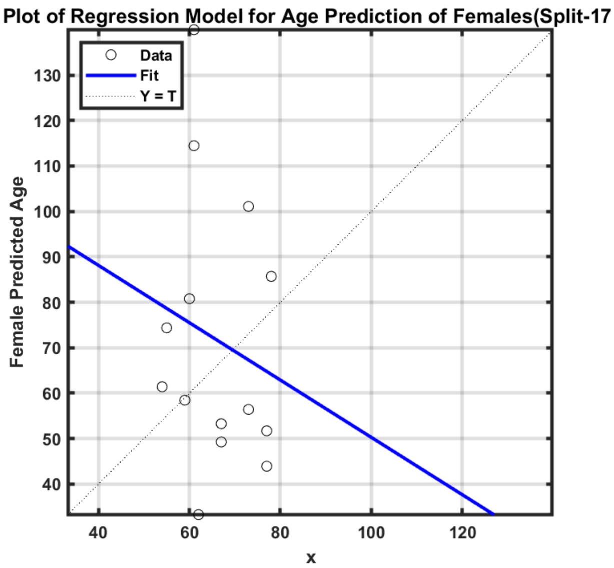

Figure 12: Regression Model for Female Age Prediction(Split-17)

4.2 Related work on the Pattern Recognition Approach

In Pattern Recognition Approach, a recent work where estimation of age is done by brain MRI images using deep learning [6]. Another approach of age estimation uses local brain features for T1-weighted images[7]. In this work, the importance of some features have been highlighted over other features and their effect in age estimation has been studied.

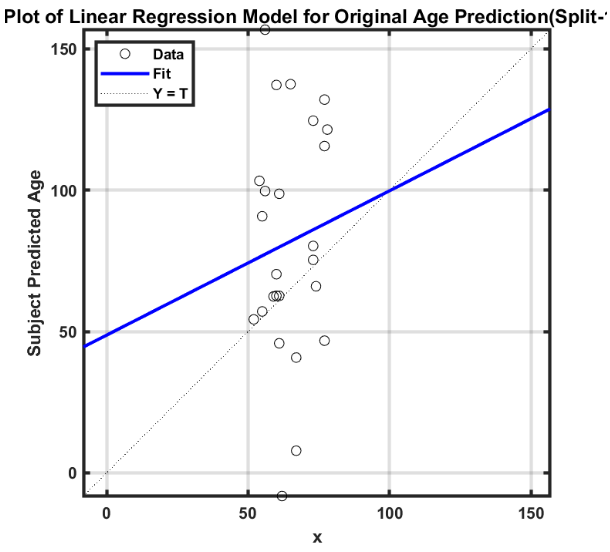

Figure 13: Original Gender Neutral Regression Model(Split-17)

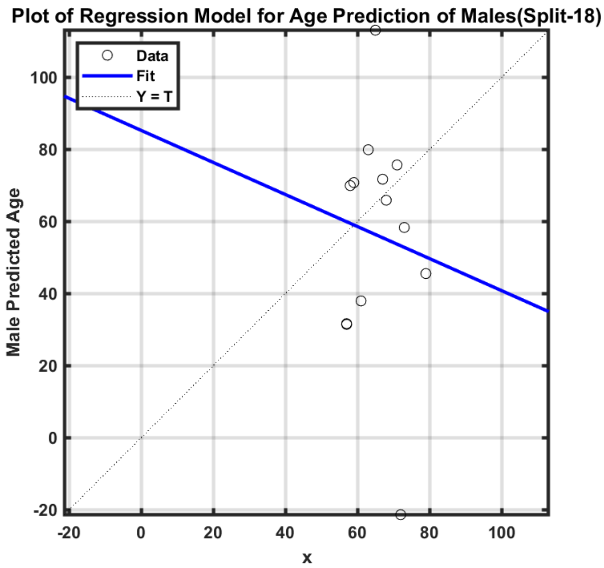

Figure 14: Regression Model for Male Age Prediction(Split-18)

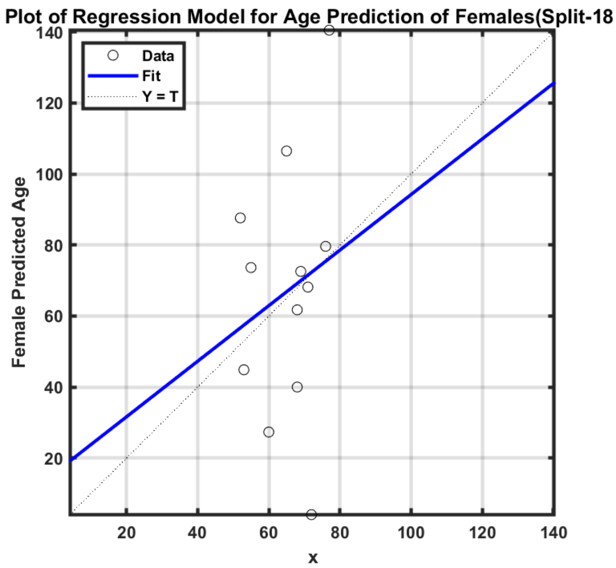

Figure 15: Regression Model for Female Age Prediction(Split-18)

5 Conclusion

Over the course of this project, we have gained superior insight on problems of feature selection and training regression and classification tasks. We have also gained evidence to comment on the two pondering questions which served at the core of this project's motivation. Based on the results which we have obtained, there are two major takeaways that we have garnered:

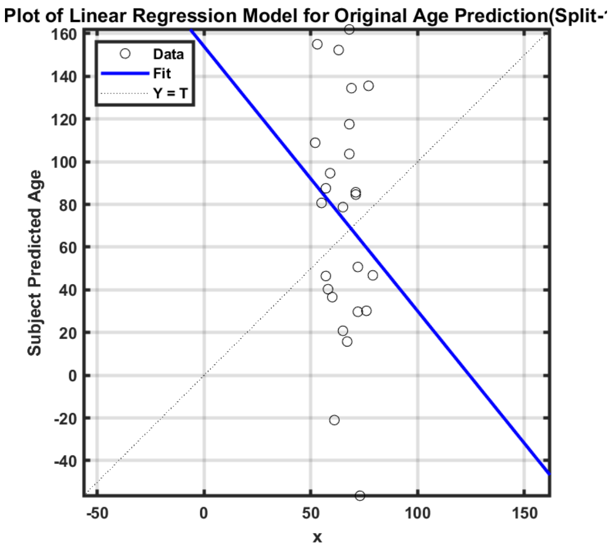

Figure 16: Original Gender Neutral Regression Model(Split-18)

- There are 248 features out of the 1000 features which can perfectly distinguish the gender of the subject. This suggests to the fact of **different asymmetry of male and female brains**.
- There is substantial difference between the output of original regression model and the separate regression models. This alludes the fact that **gender should influence age estimation**.

References

[1] L. A. Teverovskiy, J. T. Becker, O. L. Lopez, and Y. Liu, "Quantified brain asymmetry for age estimation of normal and ad/mci subjects," in *2008 5th IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, May 2008, pp. 1509–1512. 2

[2] Wikipedia contributors, "Neuroscience of sex differences — Wikipedia, the free encyclopedia," 2019, [Online; accessed 2-April-2019]. [Online]. Available: https://en.wikipedia.org/w/index.php?title=Neuroscience_of_sex_differences&oldid=890564961 2

[3] S. Baron-Cohen, *Males and females differ in specific brain structures*, 2019. [Online]. Available: https://www.cam.ac.uk/research/news/males-and-females-differ-in-specific-brain-structures 2

[4] Y. Zhang, M. Brady, and S. Smith, "Segmentation of brain mr images through a hidden markov random field model and the expectation-maximization algorithm," *IEEE Transactions on Medical Imaging*, vol. 20, no. 1, pp. 45–57, Jan 2001. 7

[5] K. Held, E. R. Kops, B. J. Krause, W. M. Wells, R. Kikinis, and H. . Muller-Gartner, "Markov random field segmentation of brain mr images," *IEEE Transactions on Medical Imaging*, vol. 16, no. 6, pp. 878–886, Dec 1997. 7

[6] T. Huang, H. Chen, R. Fujimoto, K. Ito, K. Wu, K. Sato, Y. Taki, H. Fukuda, and T. Aoki, "Age estimation from brain mri images using deep learning," in *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, April 2017, pp. 849–852. 8

[7] C. Kondo, K. Ito, K. Wu, K. Sato, Y. Taki, H. Fukuda, and T. Aoki, "An age estimation method using brain local features for t1-weighted images," in *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, Aug 2015, pp. 666–669. 8