Predict Optimized Treatment for Depression

Minakshi Mukherjee; adaboost@stanford.edu Suvasis Mukherjee; suvasism@stanford.edu CS229 Fall 2018

Objective

For the past 60 years, the anxiety and depression medications are prescribed to patients based on "The Hamilton Depression Rating Scale"(HDRS) and "Social and Functioning Assessment Scale"(SOFAS).These scores are very subjective as they are determined by clinicians based on patient interview without incorporating scientific evidence based on Brain FMRI data. Objective of this project is to identify whether HDRS score and SOFAS scores are representative of the three antidepressants prescribed based on FMRI data of 5 brain attributes.

Data Set and Features

Dataset consists of 128 patients FMRI data obtained from Williams Pan-Lab, Precision Psychiatry and Translational Neuroscience, Stanford Medicine iSPOT-D project. There are 11 features:age,gender,education,3 antidepressants: Sertraline, Venlafaxine, escitalopram, 5 FMRI brain scan data from brain region Amygdala, Insula and Nucleus Accumbens along with HDRS and SOFAS scores for all the patients. Our project analyzes both supervised and unsupervised methods. All methods are carried out independently both for HDRS and SOFAS scores. In supervised model, HDRS/SOFAS score is a dependent variable and models are fit using different combinations of brain data as feature variables. In unsupervised model, we studied 8 features(3 antidepressants) and 5 brain scan data to understand the association between medicine and brain attributes.

Models

We selected 20 patients randomly out of 128 as test set and use K-fold cross validation(K=10) on 108 patient data to train and validate our models. The following 6 supervised and unsupervised models are considered for the project.

Logistic Regression
Linear Regression
Bayesian Linear Regression with Laplace Prior
Factor Analysis
K-means Clustering
SVM

AUC for Logistic Regression

Based on Average test set misclassification error in logistic regression to predict the SOFAS logistic outcome, sensitivity and specificity of the ROC (Receiver Operating Characteristic) curve and AUC (Area under the curve) are used to understand the model performance. Based on moderately high value for AUC, we conclude that there is statistical significance between SOFAS logistic score and brain scan data.

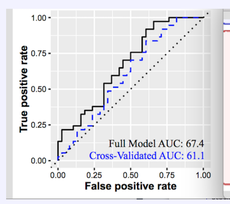


Table 1: RMSE for training and test data

Linear Regression : training= 3.4231	test= 3.6586
Ridge Regression : training=3.4244	test=3.6300
Lasso : training=3.5453	test=3.6862
Elastic Net : training= 3.5195	test=3.7181

Algorithm

Supervised Algorithm:

Bayesian Linear Regression with Laplace Prior(Elastic Net) We choose a Laplace prior for the parameter θ as Laplace distribution is symmetric around zero and it is more strongly peaked as λ grows.
The MAP estimator is the sparse lasso solution, this is useful to pinpoint the exact Brain attribute to HDRS/SOFAS score which will establish the functional connectivity between antidepressants and the specific brain region, because some of the θ's goes to zero.

$$Laplace\ Prior : p(\theta) = \frac{\lambda}{2 * \sigma} exp(-\frac{\lambda|\theta|}{\sigma})$$
$$Dataset:\ S = \{x^{(i)}, y^{(i)}\}_{i=1}^{m}$$
$$y^{(i)} = \theta^T x^{(i)} + \epsilon^{(i)}$$
$$epsilon^{(i)} \sim \mathcal{N}(\mu, \sigma^2)$$

We search for a choice of θ that minimizes the objective function

$$J(\theta) = \frac{1}{2}\sum_{i=1}^{m}(\theta^T x^{(i)} - y^{(i)})^2$$

The output of Bayesian linear regression on a new test point x_* is the posterior predictive distribution

$$p(y_*|x_*, S) = \int_\theta p(y_*|x_*, \theta)p(\theta|S)d\theta$$
$$Parameter\ Posterior\ p(\theta|S) = \frac{p(\theta)\prod_i p(y^{(i)}|x^{(i)}, \theta)}{\int_{\theta'} p(\theta)\prod_i p(y^{(i)}|x^{(i)}, \theta)d\theta'}$$

We compared Elastic net with Ridge regression and Lasso and discussed it in the "Results" section.

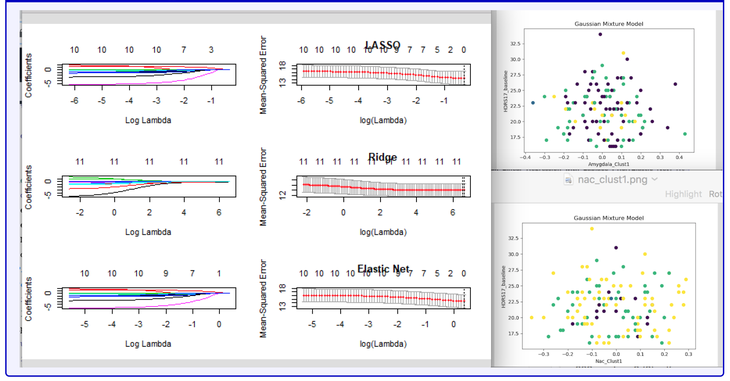
Unsupervised Algorithm:Factor Analysis

Factor Analysis works on small dataset where it helps to capture the correlations in the data.
$p(x^{(i)}, z^{(i)}) = p(x^{(i)}|z^{(i)})p(z^{(i)})$
$z \sim \mathcal{N}(0, I)$
$\epsilon \sim \mathcal{N}(0, \Psi)$
$\epsilon\ and\ z\ are\ independent.$
$x = \mu + \Lambda z + \epsilon$
$x^{(i)}\ has\ the\ covariance\ \Psi\ noise$
$\mu + \Lambda z\ is\ the\ K - dimensional\ affine\ subspace\ of R^n.$
Given the guesses for z that the E-step finds, M step estimates the unknown linearity Λ and captures the covariance $\Sigma_{x^{(i)}|z^{(i)}}$ for the posterior distribution $p(x^{(i)}|z^{(i)})$. We declare the convergence when the increase in likelihood $l(\Lambda)$ in successive iterations is smaller than the tolerance parameter. We choose the maximum of $l(\Lambda)$, out of all obtained by k-fold CV.

References

1.Estimation of Gaussian mixtures in small sample studies using l_1 penalization: https:arxiv.org/pdf/0901.4752.pdf
2.fMRI: preprocessing, classification and pattern recognition: https:arxiv.org/abs/1804.10167

Plots: Elastic Net/Lasso/ Ridge Regression and GMM for brain data



Results: Factor Analysis

In Factor Analysis, we transform the current set of variables into an equal number of variables such that each new variable is a combination of the current ones through some transformation. Here data gets transformed in the direction of each eigenvector and represent all the new variables or factors using the eigenvalues. An eigenvalue more than 1 means that the new factor explains more variance than the original variable. Output of our Factor Loadings shows that all 11 feature variables(3 antidepressants,age,gender,education,5 brain scan attributes) adequately represent the factor categories for this medical data set..

Plots: Conclusion

Based on the RMSE values and the plots above for Supervised learning, Ridge regression performs the best.Hence, HDRS and SOFAS scores statistically connect antidepressants to Brain scan data. Factor Analysis output also conforms to the same result that all 11 feature variables are important to represent the interdependent relationship among the feature variables. To get more insight, we fit Gaussian mixture model using HDRS score and Amygdala Clus 1/2 brain data as well as HDRS score and Nac Clus 1/2 brain data, however based on the above plot the representation seems unintelligible and requires further analysis.

Future Work

We would like to enhance our Gaussian Mixture Model with regtession and sparsity as follows: instead of estimating the $\mu_k\ for\ k = 1..K$, we would estimate only the coefficients of a sparse linear combinations of the X_js for all the data belonging to the same cluster using a sparsity enforcing penalty like l_1 norm of the coefficients. The main difficulty with such an approach might be to choose the right sample vector representing each cluster a priori, we would like to use Lasso as one of the potential approach to solve that problem.[1]

Link to Video Presentation

https://vimeo.com/305777481