



Predicting the Rate of Progression of the ALS Disease in Patients

Durga Ganesh, Michael Black, Neeharika Madduri
Stanford University



Abstract

ALS is a neurodegenerative disease with few known causes. The rate of progression of the disease varies significantly as some patients live many years with the disease while others succumb to the disease within a couple of years.

Our goal was to try to predict the rate of progression of the disease in patients given a collection of predictors taken during the first three months of a clinical trial.

We first used some feature selection methods to reduce the number of features that we would consider. We then employed some regression prediction techniques that worked well with nonlinear data -- namely decision trees, random forests, and neural networks.

We also broke the data up into two groups (fast decline & slow decline) and used the SVM to try to predict the category the decline of the disease would fall in. With both the category and regression prediction, we could be more certain of a given prediction.

Feature Selection

Initial data set had 858 features with 2424 patient records. Missing values were replaced with the median of that feature.

As a first pass, to reduce the number of features, we performed L-1 regularization. The L-1 regularization cost function is shown below:

$$\sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 + \lambda \sum_{j=1}^p |\beta_j|$$

This regularization problem can also be formulated as an optimization problem, where the coefficient estimates solve

$$\underset{\beta}{\text{minimize}} \left\{ \sum_{i=1}^n \left(y_i - \beta_0 - \sum_{j=1}^p \beta_j x_{ij} \right)^2 \right\}, \quad \text{subject to} \quad \sum_{j=1}^p |\beta_j| \leq s$$

where s can be interpreted as the budget you have for the magnitude of your coefficients. As we tune this budget value, only the coefficients corresponding to the most relevant features remain. We could reduce the number of features to 34 after L-1 regularization.

After running least squares with L-1 regularization, we got a mean squared error of 0.3795 which is an improvement over the strictly linear model.

We used cross validation to find the best lambda. Figure 1 shows the mean squared error (and its standard error) for different values of lambda evaluated on our test set.

Figure 2 shows how the coefficients change as we increase the budget s shown in the equation above. As we increase the budget, more coefficients can be non-zero which approaches the least squares solution.

We further implemented Forward and Backward Selection. After using these methods, we were down to 23 features which gave us the best prediction results.

Lambda Cross Validation

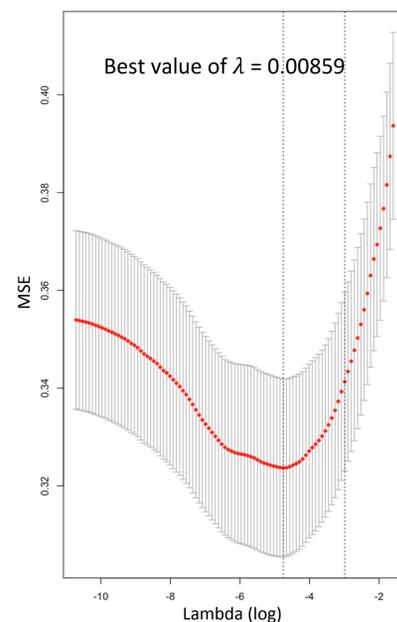


Fig 1. CV Error for lambda tuning in L-1 regularization

L-1 Regularization Coefficient Size

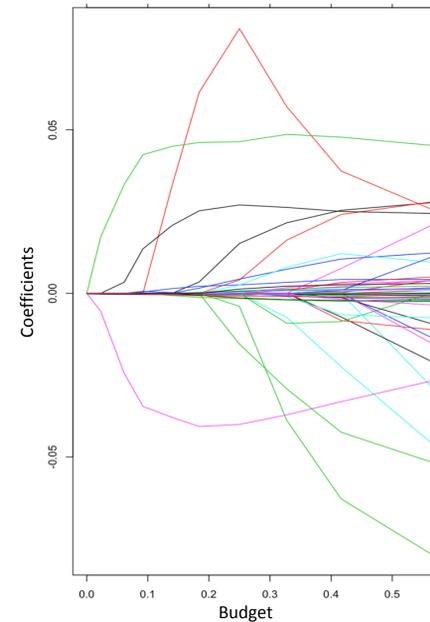


Fig 2. Coefficients as we vary L-1 regularization budget

Table 1. Test MSE for various methods

Method	Test MSE
Predict Mean	0.4284029
Linear Regression	0.5567662
Lasso	0.3795065
Neural Networks	0.3385600
Random Forest	0.3364572
SVM Regression	0.3389020
Boosted Decision Tree	0.3304962

Random Forest MSE

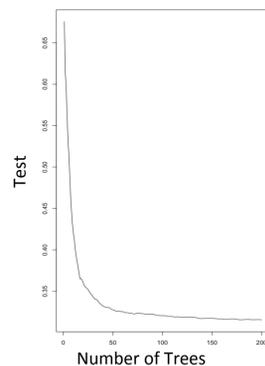


Fig 3. Test MSE as we vary number of trees

Boosted Decision Tree MSE

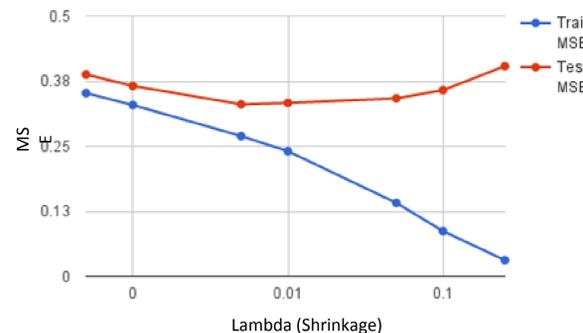


Fig 4. Training & Test MSE for shrinkage parameter with fixed no. of trees

Test MSE vs number of neurons

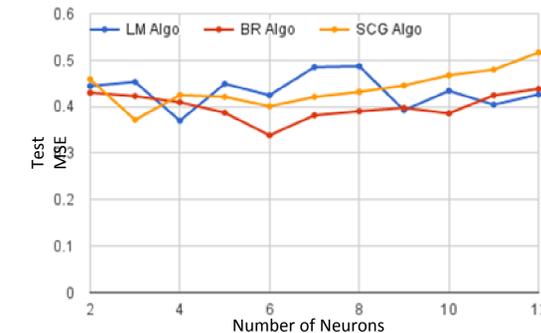


Fig 5. Test MSE as we vary the number of neurons in hidden layer

Best Validation performance is 0.305

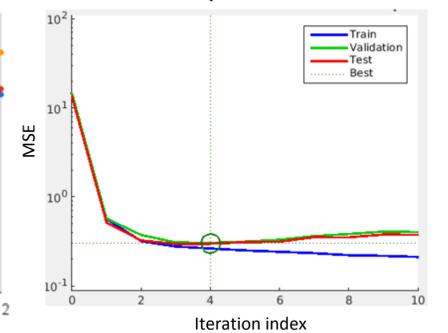


Fig 6. MSE for train, validation, & test sets

Results & Discussion

We used four main regression techniques to predict the rate of decline of the ALS disease in patients. The test MSE of the methods is summarized in Table 1.

Neural Networks

We trained our neural network model with three different algorithms - Levenberg-Marquardt, Bayesian Regularization and Scaled Conjugate Gradient on models with different number of hidden neurons. The test MSE for all the combinations is shown in Figure 5.

The best model for the data has 6 hidden neurons and is trained by Bayesian Regularization Algorithm as shown in Figure 6.

SVM Regression with Gaussian Kernels

We used SVM as a regression technique & cross-validation to select the cost parameter in the SVM, in order to get the best prediction.

Decision Trees

Decision trees aim to split the features into subgroups and then predict a piecewise constant value at the final leaves of the tree. We tried a couple of techniques that would reduce the variance of our procedure. One such method was with a **Random Forest** with 1000 trees and 5 features considered at each split.

Another technique was a **Boosted Decision Tree**. The rate at which the method "learns" is controlled by the shrinkage parameter. Figure 4 shows the training and test MSE found using cross validation.

We also categorized the data into two categories -- fast and slow progression of the disease, to transform the problem into a classification setting. We implemented this using an SVM with a Gaussian kernel, and achieved a classification accuracy of 99.3%.

Conclusion

Overall, the data was very nonlinear and didn't show many obvious patterns that we could exploit. Through feature selection and transforming the data to higher dimensions using kernels we were able to make good predictions.

References

G. James, D. Witten, T. Hastie, R. Tibshirani, An Introduction to Statistical Learning with Applications in R, Springer, New York, ISBN: 978-1461471370, 2013.

Contact

Durga Ganesh: durga@stanford.edu
Michael Black: mblack1@stanford.edu
Neeharika Madduri: nmadduri@Stanford.edu