Final Project
Stat 215A, Fall 2017

Alexander Brandt

December 10, 2017

Introduction

In this lab we take a look at fMRI data from human subjects exposed to various images in an attempt to understand the neurological activity of the visual cortex. The activity is determined based on regions of the brain which have been partitioned into subset voxel regions. An initial goal of this lab is to try and establish some form of coherent model between the fMRI measurements, which have been transformed to an ensemble of features via the Gabor transformation. The response to the images is determined as a univariate measurement.

Exploratory Data Analysis

First we attempt to understand the voxels themselves in the context of their spatial patterning and similarity. We begin by correlating the responses of all voxels with one another and plotting them, and also specify an additional hierarchical clustering of the correlations. There are two fairly prominent clusters, which we box off in the heatmap. Then we plot these voxels based on their x, y, and z locations in the brain, colored by the heatmap cluster identity. Please note that this is a screen shot from the plotly rendering, which is provided as an HTML file in the figures directory. The first cluster, comprising V1 through V9, and V12, V14, V15, V16 through V19. The second cluster comprises V10, V11, V13, V16, and V20. We see that the first cluster comprises the interior of the visual cortex, and the second cluster comprises the exterior of the visual cortex. It should be noted that the interior comprises a series of highly and internally correlated

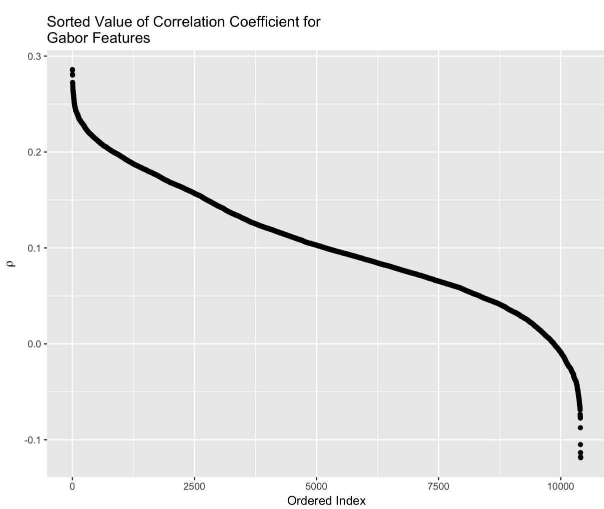

Figure 1: Overall per Feature Correlation

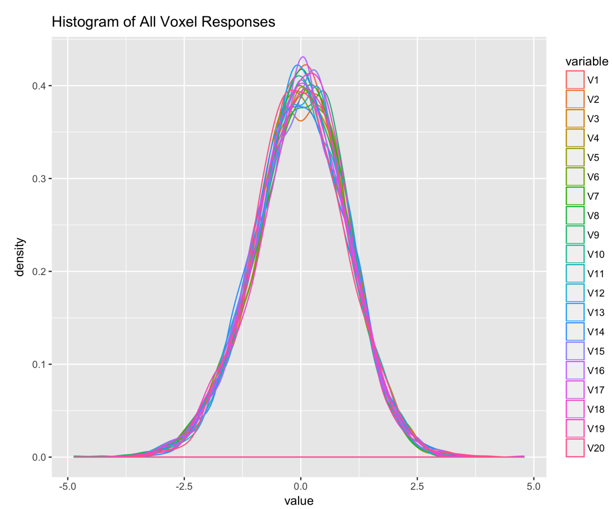

Figure 2: Response Distribution per Voxel

voxels, and the outer voxels seem to bear little similarity with either themselves or the voxels in cluster 1. This partitioning will be useful in understanding our lasso and ridge regression models later in the lab.

1 Partitioning the Data

Gleaning some wisdom from Elements of Statistical Learning by Trevor Hastie, a good rule of thumb for data partitioning can be 50% of the data for training, 25% of the data for validation, and 25% of the data for testing. We also have an additional testing set has been given for us as part of the lab's assessment. Though this won't be relevant until the very end of the lab. We will follow Hastie's partitioning scheme for our lab.

To partition, we randomize the indices and subset them into 1/2, 1/4 and 1/4 portions (rounding to the nearest whole number).

2 Training LASSO and our Linear Models

- Cross Validation (CV) (Package Installation)
  - Advantages: A robust method for selecting a correct λ coefficient based on multiple trials with the training data. Literature seems to mention it very favorably.
  - Disadvantages: Conditions on MSE rather than our desired parameter (correlation with measurements). Slower than other methods since it must be run k (in our case, k=10) times.
- Cross Validation (CV) (Coded Based on Correlation)
  - Advantages: Attempts to condition the selection on the value most important (correlation between predicted value), rather than MSE in the glmnet package automated CV function.

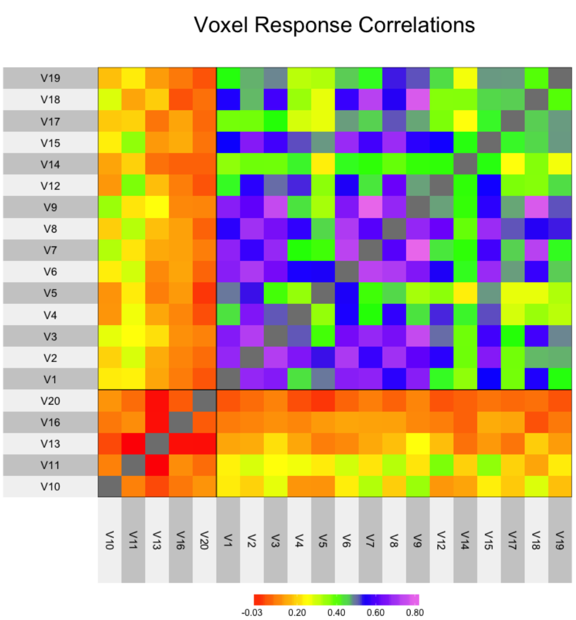

Figure 3: Voxel/Voxel Response Heatmap Correlation

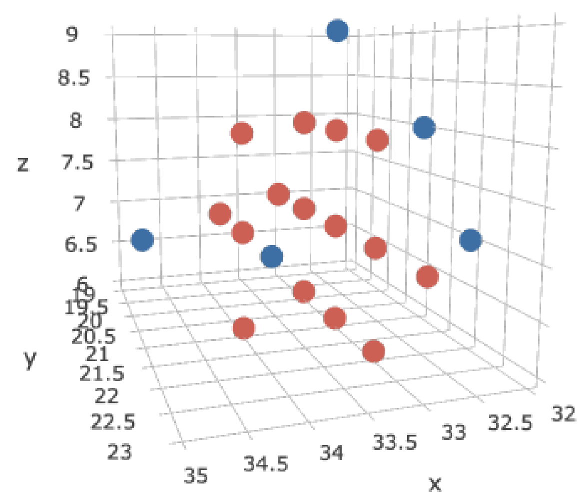

Figure 4: Voxel 3D Orientation and Cluster Membership

  - Disadvantages: Slower than other methods since it must be run k (in our case, k=10) times. Selects more features than necessary to get a similar correlation when compared with ES-CV.

- Estimation Stability with Cross Validation (ES-CV)
  - Advantages: Selects less features overall when compared to standard CV. Therefore it finds the most parsimonious form of the model.
  - Disadvantages: Doesn't dramatically outperform CV with respect to correlation coefficient.

- Akaike Information Criterea (AIC)
  - Advantages: Fairly conservative with respect to feature selection (it selects more features), and does well within the ensemble of models provided to it.
  - Disadvantages: Provides no information about the absolute quality of the model, only the performance relative to other models.

- Akaike Information Criterea with Finite Sample Size Correction (AICc)
  - Advantages: All the advantages of AIC with an additional term that emphasizes the difficulties of an underpowered data set.
  - Disadvantages: Provides no information about the absolute quality of the model, only the performance relative to other models. Functional form can garner discontinuities with wider ranges of lambda.

- Bayesian Information Criterea (BIC)
  - Advantages: More appropriate for finding a true model in the candidate set. More likely to select a parsimonious model.
  - Disadvantages: Not assymptotically optimal (compared to AIC, which is). It is often rare that the "true" model is in the data set provided to it.

- Standard Linear Model
  - Advantages: Few, if any in these contexts. A good baseline for seeing the improvements that ridge and lasso bring.
  - Disadvantages: Weights all components equally. Probably would perform better if the features were limited to the first 500-1000 features (with respect to correlation between the features and the voxel responses).

Each training set produced several different values λ in the case of both the lasso and ridge regression (See these criterea in figure 5).

3 Model Performance

Given that the Gallant lab uses correlation between the fitted and observed values for the asessment of their models, we will use this criteria with a validation set in order to assess which model performs the best. From Figure 5 we see a wide spread of all possible methods, with a notable exception of the ridge regression (though as we shall see, ridge does not uniformly out perform lasso). Also interesting to note is that when we take our best model and run it on our remaining test data, most correlations are centered at about $\rho = .45$ with some outlines. Upon closer inspection, the lowest performing voxels (V10, V11, V13, V16, and V20) all correspond to the second voxel cluster of poorly correlated features. They seem recalcitrant to our modeling efforts, and thus will be the subject of less focus in the report.

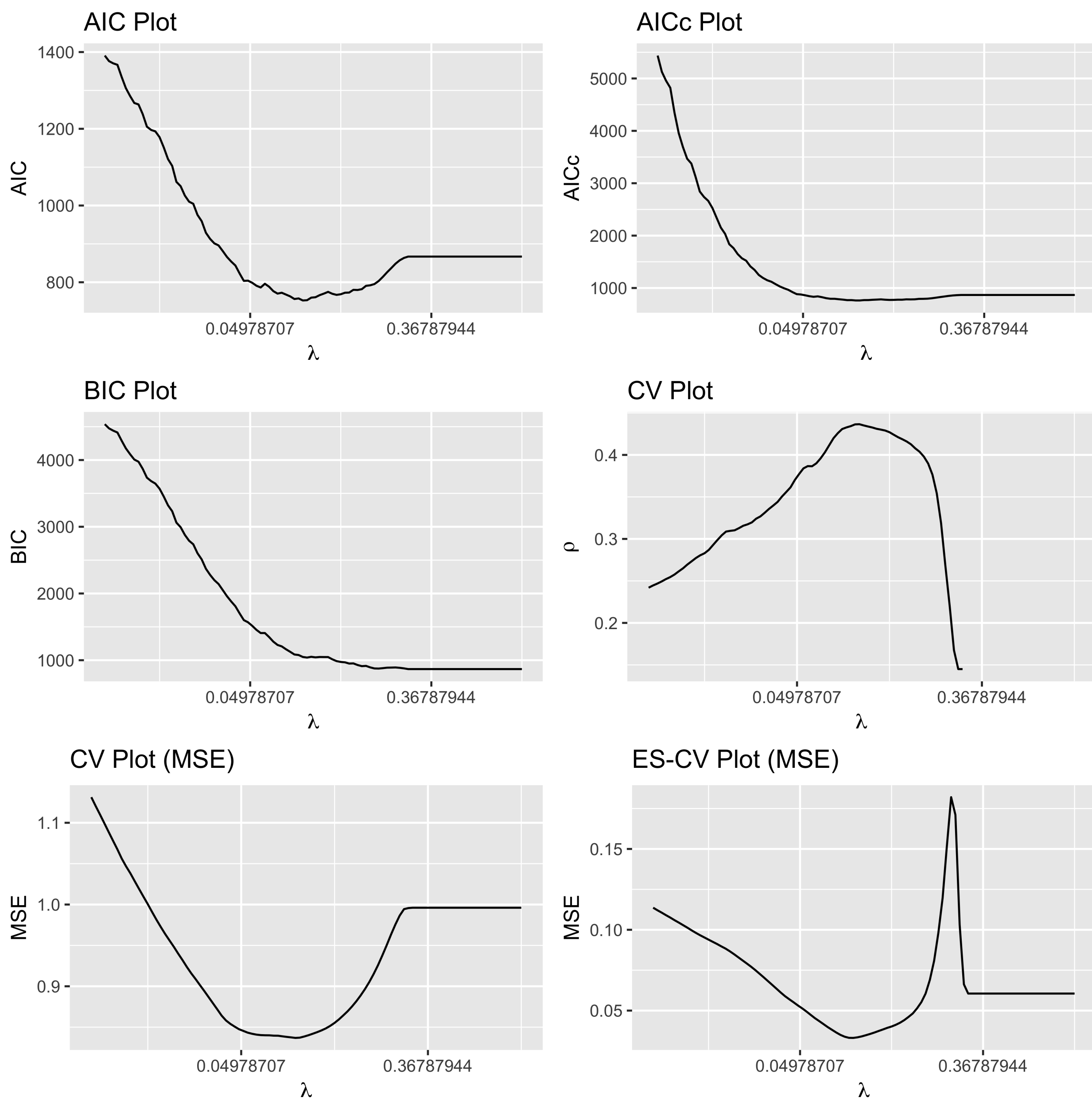

Figure 5: Minimized/Maximized Metrics for Choosing λ

Voxel	Test Set Correlation	Model Selection Criteria
V1	0.46013295	ES-CV
V2	0.52738720	ES-CV
V3	0.45265275	Package CV
V4	0.47959658	ES-CV
V5	0.50333942	AIC
V6	0.46547010	CV
V7	0.51441616	ES-CV
V8	0.53194851	AIC
V9	0.54876383	Package CV
V10	0.16325771	BIC
V11	0.22881389	ES-CV
V12	0.47577744	AIC
V13	0.26678541	AIC
V14	0.29745503	AIC
V15	0.52170079	AIC
V16	0.07446250	AIC
V17	0.35691117	AIC
V18	0.42511160	Ridge
V19	0.34618609	ES-CV
V20	0.01045314	ES-CV

4 Diagnostics

4.1 Model Fit and Outliers

We exclude no observations from our model, given that it has been heavily curated by the GSI/professor. If we were to identify outliers, we would use a multivariate outlier detection method based on robust methods. This a method generally based on normality assumptions (which we can largely confirm given our EDA analysis). It uses a multivariate Mahalanobis distance to determine outliers [2]. This method identifies 60 images (3.42%), one of which will be discussed later. The image indices that are identified are:

```
150 164 194 211 262 289 315 375 389 405 427 430 442 445 455 464 493 519
531 540 547 567 574 597 615 645 663 686 731 783 804 833 836 838 924 935
938 942 954 963 990 1002 1006 1011 1029 1041 1042 1047 1199 1341 1430
1461 1531 1539 1605 1645 1654 1662 1696 1717.
```

One of these images will show up in our interpretation set. It favors our models to exclude it given the relevance of the other images.

4.2 Stability of Prediction

We will restrict our analysis of stability to the first voxel (V1) and investigate it further with bootstraps in section 5.2.

5 Model Interpretation

5.1 Feature Sharing Between Models

Only 592 (¡5%) features were selected for all of the top models for every voxels. We report the top 13 (all found in a 4 or more of the models) in figure 8. Cleary there is a prioritization on a very small subset of the gabor features in determining image response in our models.

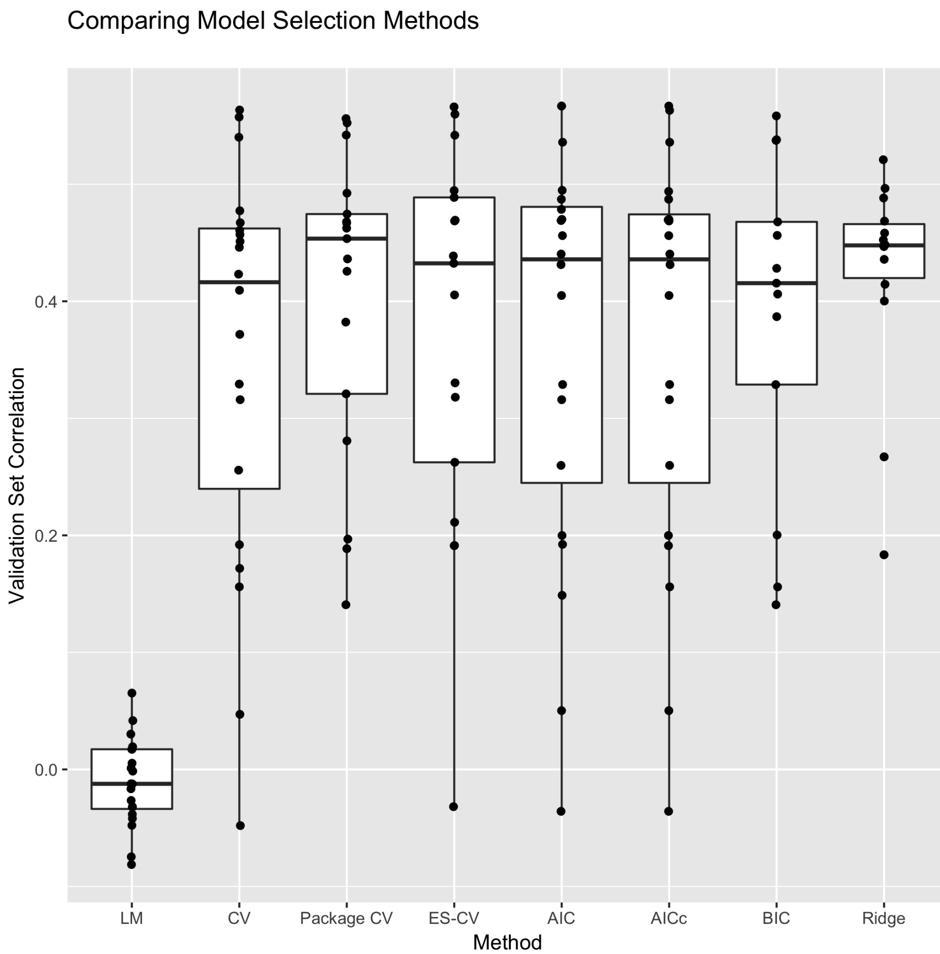

Figure 6: The Range of Validation Set Correlation During Model Selection

5.2 Correlation Stability Across Bootstraps

For Voxel 1 we sample with replacement from our test set 1000 times to create 1000 bootstrap data sets (resulting in both a design matrix as well as a response matrix/vector). These are shown in figure 9. We then predict from our best model, which was a lasso regression chosen by ES-CV, and check its correlation with the response matrix/vector. The results are shown in figure 7, which is a histogram of the 1000 bootstrap correlation values. We see that our model is fairly robust with respect to the true test set correlation coefficient, ($\rho = 0.46013295$). The distribution is roughly centered around this point and shows a reasonable spread around it ($\sigma = 0.03682339$).

5.3 Hypothesis Testing

To use our models in various hypothesis testing schemes, we would need to extract the test statistic. Though there is no readily available translation for this test statistic into a p-value, we could certainly compare various features to one another in terms of significance to their contribution to the prediction of voxel response. There are certainly proposed test statistics, though there is not a readily available rule. The bootstrap method will also allow us to compute a confidence interval that we could use to perform hypothesis testing.

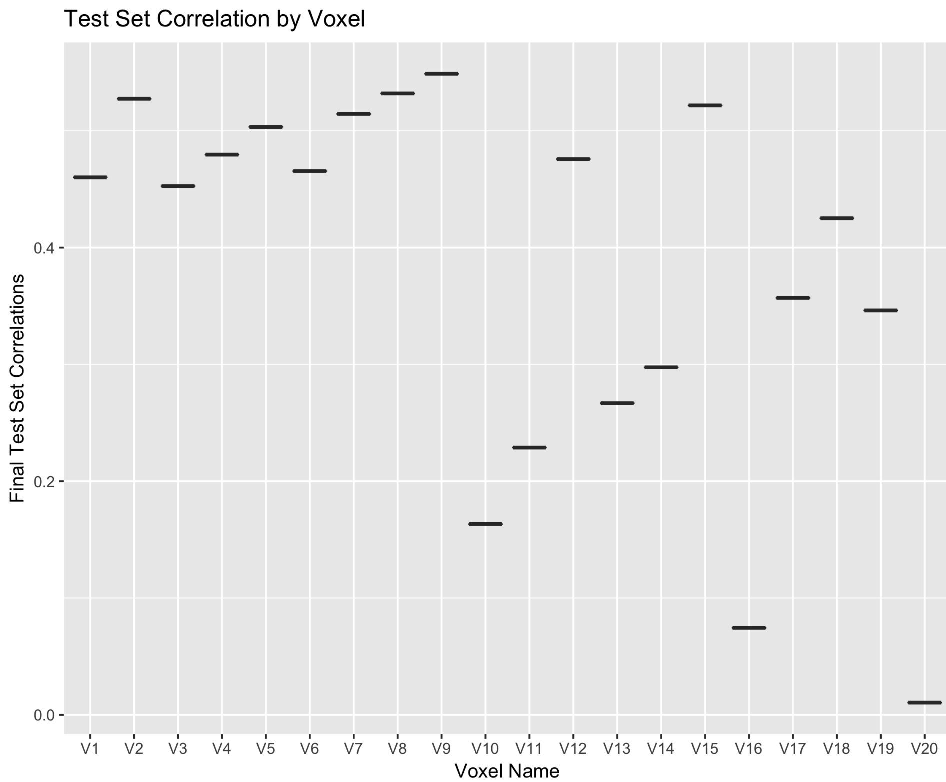

Figure 7: Examining the Test Set Correlation for the Best Model by Voxel

5.4 Understanding Voxel Response

Given that our model of V2 is one of the higher performing models, we will see which images specifically in our test set generate the strongest response. We take the four images with the strongest predicted response and display those images in a grid in figure 10. I was pleasantly surprised that the images all seem to display animals, with the exception of the statue face. Though this is certainly a non-living humanoid picture, it is also identified as an outlier in our outlier analysis. This might allow us to forgive its inclusion. Either way, the intuitive cogency of the image content is hopefully an indicator that our modeling has been marginally successful.

6 Predictions

Given our model performance, we use the LASSO model selected with ES-CV to predict the responses to the 120 testing images for the first voxel (V1). It is stored in `data/predv1_AlexanderBrandt.txt` file in the project directory.

References

[1] Yanjun Wang, Qun Liu, Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stockrecruitment relationships, In Fisheries Research, Volume 77, Issue 2, 2006, Pages 220-225, ISSN 0165-7836, https://doi.org/10.1016/j.fishres.2005.08.011.

[2] Elena Gimnez, Mattia Crespi, M. Selmira Garrido, Antonio J. Gil, Multivariate outlier detection based on robust computation of Mahalanobis distances. Application to positioning assisted by RTK GNSS

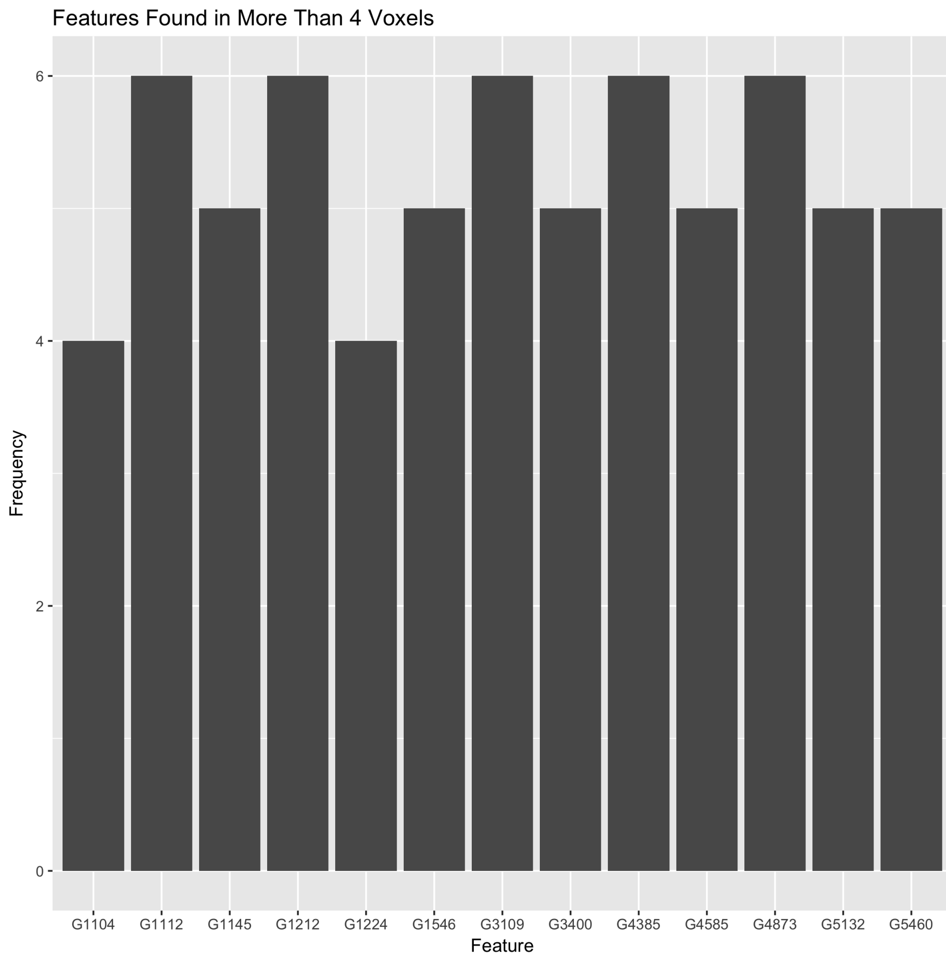

Figure 8: Distribution of each feature for two types of errors

Networks, In International Journal of Applied Earth Observation and Geoinformation, Volume 16, 2012, Pages 94-100, ISSN 0303-2434, https://doi.org/10.1016/j.jag.2011.11.011.

Acknowledgements

The author greatly appreciated the help of Max Gardner when discussing various parts of the lab, especially the model selection metrics.

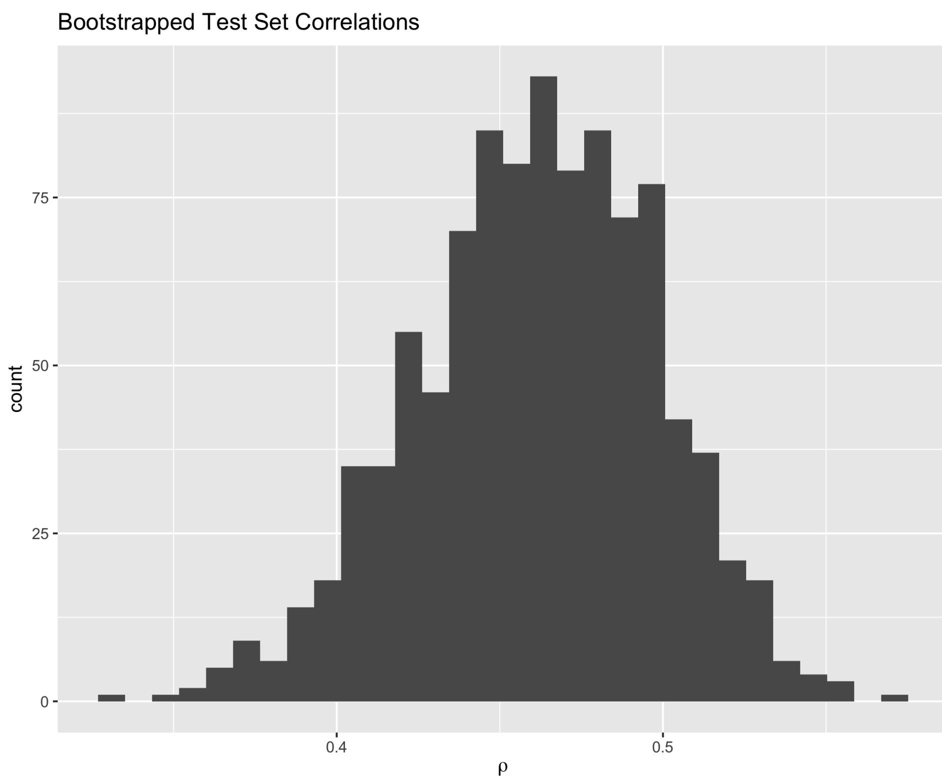

Figure 9: Histogram of bootstrap correlations

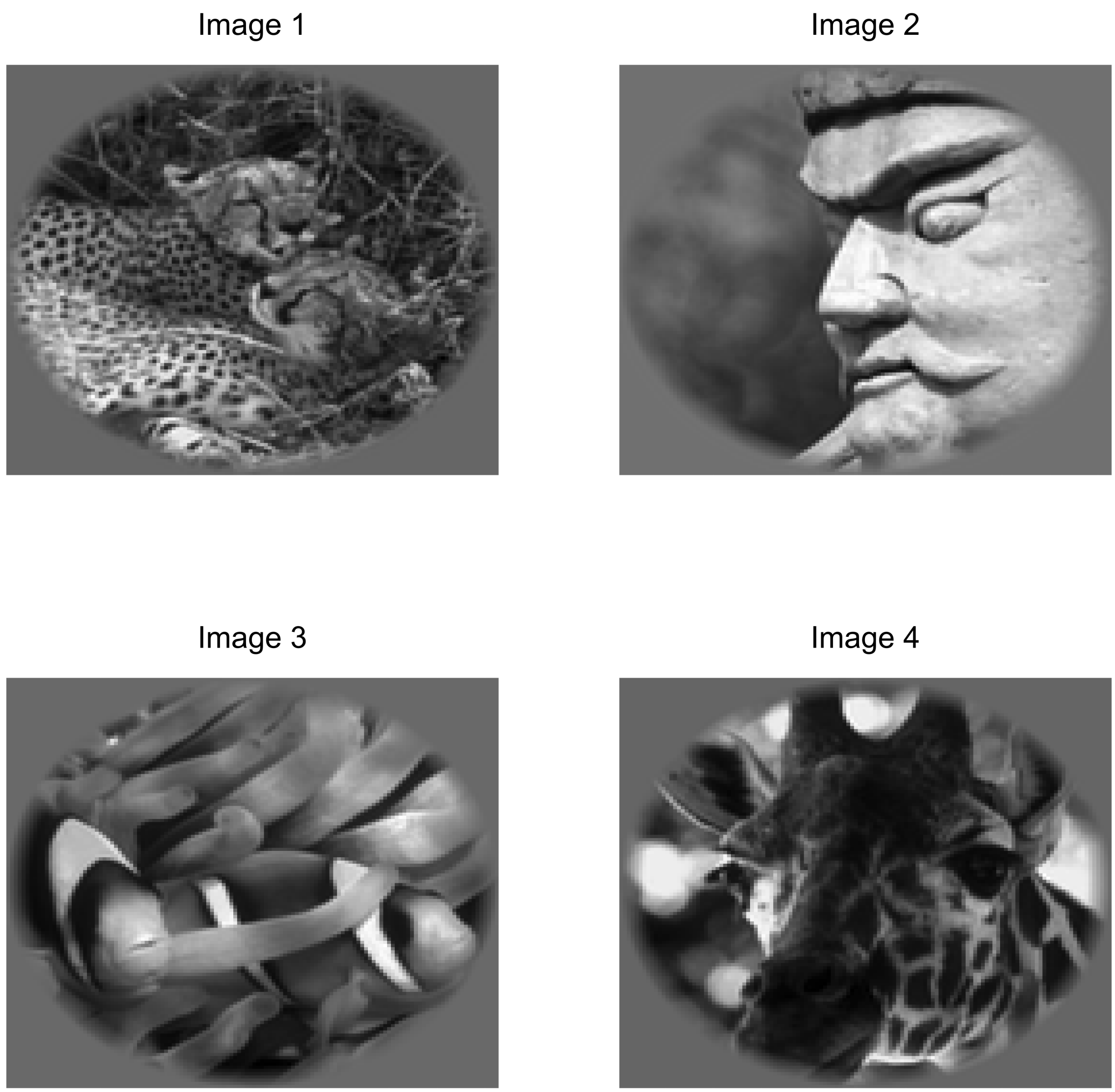

Figure 10: Top Response Images for Voxel 2