Predicting Adolescent Gender Attitudes Using NLSY

Shweta Kamath **Nivedita Vatsa** **Carolyn Vilter**

1 Research Question

We proposed to predict adolescents' attitudes towards gender roles by applying machine learning methods to the detailed information available in the U.S. National Longitudinal Survey of Youth (NLSY). The longstanding survey program asks both mothers and children about their views on gender roles along with a wide range of other demographic, economic, and social questions. Though a subjective concept like "gender attitudes" presents descriptive and predictive challenges, the NLSY's depth and breadth is able to provide an exceptional level of information about children's lives and, by extension, some insight into their perspective on gender.

2 Data

Our analysis draws on two surveys under the banner of NLSY, a program initiated by the Bureau of Labor Statistics in 1979. The first survey, the NLSY-79, samples a cohort of young people initially between the ages of 14 and 22 and follows them over the course of their lives. NLSY-79 data is available annually from 1979 to 1992, after which it is available biennially.

The second survey from which we sourced data is the NLSY-YA (Children and Young Adults). Starting in 1986, the children of female respondents in the original NLSY-79 cohort were sampled under the NLSY-YA, also on an ongoing longitudinal basis. The data resulting from both survey programs is publicly available, and children surveyed under the NLSY-YA are linked to the unique ID of their mothers in the NLSY-79.

The NLSY-79 and the NLSY-YA together present a unique opportunity for research because they allow us to link a rich dataset on children to equally detailed information on their mothers over many years.

3 Literature Review

Our research question and approach to shaping and analyzing the abundance of NLSY data were informed by our literature review. Since our project lies at the intersection of many disciplines, we focused on two (overlapping) groups of existing literature: research on gender and research that applies machine learning methods to the NLSY.

3.1 Gender Studies

"Gender attitudes" refers to individuals' views regarding the roles men and women should play in society. Strong gender stereotypes can restrict people's - especially women's - expectations of their abilities. Stereotypes also narrow the universe of opportunities available to members of a society. Developmental researchers have identified that rudimentary stereotypes start forming in children around the age of 3 and gender stereotypes start solidifying around ages 11-14 (GEA Study).

There is a large body of research that tries to understand what factors influence egalitarian beliefs in adults. Studies looking at the lifetime of an individual that track the change in beliefs over time find that marriage, hours worked, and parenthood influence beliefs. Amongst the most studied factors is education, which has a positive effect on egalitarian beliefs. There is also evidence supporting the fact

that increased levels of religious practice reinforce traditional gender roles and reduce egalitarianism (Hertel and Hughes 1987, Peek *et al* 1991). An important way to understand the beliefs that a child develops is to look at the environment that they grow up in. Sutfin *et al* find that intergenerational transmission of beliefs occurs through direct interaction, modeling, and the construction of the child's home environment.

Informed by this literature, we carefully selected the NLSY questions and corresponding features that we felt were most likely to be associated with a child's perspective on gender.

3.2 Applications of Machine Learning to the NLSY

Our approach was also informed by the narrow but highly relevant literature applying quantitative, machine-learning methods to the NLSY. Lukes (2018) uses NLSY data to estimate a series of models that evaluate the long-term impact of the Head Start Program. The paper first estimates a least squares model. It also deploys a Lasso model specifically to combat the high-dimensional nature of the NLSY and identify the most relevant regressors. The regularization parameter is selected by trial and error. From LS to Lasso, he finds that the effect of participating in Head Start increases in magnitude, but since the overall magnitude is still small, he concludes that it has no meaningful long-term impact.

In "Uncovering Sociological Effect Heterogeneity Using Tree-Based Machine Learning," Brand *et al* use decision trees to identify NLSY population subgroups with different responses to social interventions. The paper describes the importance of moving beyond sociology's demographic defaults, like race and gender, toward more sophisticated analyses of the dimensions along which a population displays disparities and variation. The discussion supports the idea that traditionally qualitative topics - like our research question - stand to benefit from the application of machine learning methods.

4 Data Cleaning and Preparation

Preparing the convoluted NLSY data for analysis required a substantial investment of time to extract, clean, and reshape the data. NLSY-79 data, capturing mothers' data, is available for 28 rounds. The NLSY-YA, capturing the children's data, has 17 rounds for the child sample. Each round of the survey contains thousands of variables exploring demographics, attitudes and behaviors, educational attainment and aptitude, family environment and many other topics. The data also varies between years in the language of the questions asked and the answer options provided. Representing any single concept - or even any single survey question - in our final, combined dataset required several layers of conceptually- and technically-informed decision making.

4.1 Feature Selection

Our outcome of interest was children's gender attitudes during adolescence, which we defined as the latest available survey year when a given respondent was age 11-14 (a moment in time referred to henceforth as "time of Y"). The NLSY captures gender attitudes for children and mothers via five questions answered using a Likert scale. Prioritizing questions according to data availability, we chose to focus on the survey item "Girls and boys should be treated the same at school."

We selected feature variables to provide relevant information about children's and mothers' characteristics and experiences. We drew guidance from a study conducted using NLSY data on adolescents' gender ideologies which emphasized demographics, religion, mother's employment, father's involvement, etc. Our dataset captured these features along with basic demographic characteristics of the mother and child, the poverty status of the household the child grew up in, the responses of the mother to gender attitude survey questions, and other relevant features (see Appendix for complete list).

4.2 Feature Cleaning and Manipulation

We first converted the longitudinal data into a child-level cross-section by identifying each child's "time of Y" survey year and extracting information in relation to that year.

The next cleaning steps included creating a binary version of the Y variable for experimental purposes, one-hot-encoding our categorical feature variables, aggregating dummy categories that were too

small to be meaningful alone, and standardizing our features to account for differences in magnitude (between, for example, birth years and 0/1 binary variables). We chose not to normalize the data because we did not have any *a priori* assumptions about the distributions underlying the data.

The data contain various kinds of missing values, *e.g.,* non-response, uncertainty, *etc*. We chose to drop observations with missing values, though we recognize that since the data are likely not missing at random, this may bias our final results. In its final form, the dataset has 5,161 rows, 2 dependent variables, and 43 features.

5 Methods

5.1 Least Squares

We implemented a least squares approach as a baseline model. Our first analytical question was whether least squares would perform better when predicting a binary representation or the original 4-category representation of our dependent variable, children's Likert scale responses to the statement "Girls and boys should be treated the same at school." We created a version *y-scale* which mirrored the original format of the survey question (1 = Strongly Agree, 2 = Agree, 3 = Disagree, 4 = Strongly Disagree) and a second version *y-binary* which coded agreement as 1 and disagreement as -1.

For each version of the model, we ran 100 rounds of cross validation. Each round randomly divided the 5,161-row dataset into 10 approximately equal-sized sets: 9 constituted the training set and the remaining 1 the test set. Our function returned the average error over 100 rounds, the $\hat{w}$ array associated with the lowest-error iteration of the model, and plots of the distribution of predicted labels vs. true labels as well as a histogram of test error over all iterations. This allowed us to estimate the accuracy of our least squares model and effectively compare the *y-scale* and *y-binary* implementations.

5.2 Lasso

Since our current framework includes 43 features related to education, religion, income, *etc.*, we decided that a Lasso model would be suitable for selecting the most relevant features. The Lasso model employs L1 regularization. In other words, the objective function to minimize can be expressed as

$$\Sigma_{i=1}^{m}(y_i - \Sigma_{j=0}^{p} w_j x_{ij})^2 + \lambda \Sigma_{j=1}^{p} |w_j| \tag{1}$$

where m is number of observations, p is the number of features, and λ is the penalty on the size of the weights, w. Unlike the ridge regression (as seen in class), there exists no closed form solution for this problem because the second term is not smooth and differentiable. Therefore, in order to solve this problem, we implement an extension of gradient descent called the iterative shrinkage-thresholding (ISTA) algorithm and subsequently, a faster version of this algorithm called FISTA. ISTA tackles the smoothness issue by separating the objective function into smooth and non-smooth components, and then at each iteration, presents the non-smooth component in a differentiable form. FISTA builds upon this by modifying the iteration process so that each update to the weight depends on the previous iteration (also termed as "accelaration"). The code for these algorithms, which we have annotated, is provided in the Appendix Lasso Code Snippet.

Repeating the cross-validation strategy from our least squares analyses, we implemented both the ISTA and FISTA algorithms to again predict gender attitude responses based on our features. We used this approach on a variety of λ values and selected the "best" (λ^*) as the parameter corresponding to the lowest cross-validated test error error rate. Our starting values for the w vector was zero.

6 Findings

6.1 Least Squares

Through 100 rounds of cross validation, we estimated that our least squares model for predicting the 4-category *y-scale* representation of gender attitudes had an error rate of approximately 61 percent, where error is defined as the percentage of predicted y labels that did not match the true y label. The error was roughly normally distributed around 61 percent, with instances of much lower (minimum

32 percent) and much higher predictive error in a given round (Appendix Figure 1). Given that our variable of interest had four categories, an error rate of 61 percent means that the least squares model could assign gender attitude survey responses to children with significantly greater-than-random accuracy (less than 75 percent error).

We noted that the model successfully assigned every possible label in its predictions - 1, 2, 3, and 4 - despite the overwhelming dominance of the "1" label in the data. Further work would be required to rectify the distribution of the label predictions, which is inconsistent with the true distribution for the labels 2, 3, and 4 (Appendix Figures 2, 3).

We also implemented least squares for the second variation of our y variable, *y-binary*, which assigned 1 to all levels of agreement with "Girls and boys should be treated the same at school" and -1 to disagreement. We found that the model predicted *y-binary* with a mean error rate of 58 percent: slightly, but not significantly, better accuracy than *y-scale*. Given the much wider spread of *y-binary* test error over 100 iterations (Appendix Figure 4) and the greater detail captured by the *y-scale* variable, we elected to proceed with predicting only *y-scale* in the Lasso model.

6.2 Lasso

The lasso model allowed us to examine our features and determine which ones were more significant. As expected, the number of non-zero parameters falls as λ increases (Appendix Figure 5). This is also visualized in Figure 6, showing all features (green) for smaller λ, but only some features (pink) for larger lambda.

Through 100 rounds of cross-validation, we find that the Lasso model performs progressively worse for larger values of λ, implying that the least squares model is in fact the preferred approach for this problem. Indeed, for a λ as low as 0.1, we observe an error rate of 86 percent. This generates a cross-validation training error of 86 percent (Appendix Figure 7).

Although the model does not perform well, it might help us shed light on relevant features. The Lasso model with $\lambda = 0.1$ penalized the following coefficients down to zero: year at Y, whether mother's religion falls in an "other" category, and response to the statement that a mother working leads to juvenile delinquency in her children. It is interesting that year was not a strong predictor, suggesting that association between X and Y is similar across time regardless of whether a child was surveyed in the 1980s, 1990s, or 2000s. The model found that the strongest predictors of gender attitudes were a child's confidence index (higher confidence associated with more egalitarian views) and expected education attainment (higher expected attainment associated with more egalitarian views). Although the Lasso model had an unimpressive error rate, it did provide helpful insight into determining which features had more or less explanatory power.

7 Conclusion

Our process illustrated that finding, understanding, and preparing data are time-consuming and important parts of the machine learning process. We learned that some datasets and features may lend themselves better to a quantitative analysis than others. For example, our central focus, "gender attitudes," is an abstract and complicated idea that proved difficult to predict with high accuracy. Still, we were pleased to develop a least squares model that achieved significantly better-than-random predictions. We also found the Lasso model to be useful in confirming our hypothesis that education and self-confidence are important predictors of gender attitudes.

Future work on this topic would need to explore different representations of "gender attitudes" along with different features, new feature engineering approaches, and alternative functional forms. For example, there is a need to better understand the role played by the father in the child's upbringing. Our analysis also suggested that some variables do not have a linear relationship with the outcomes, including education and hours worked.

Above all, our research suggested that doing justice to the topic of children's gender attitudes and their origins would require significantly more time, issue area expertise, and attention to model specification and tuning. But it's clear from our findings that a comprehensive and detailed dataset like the NLSY, coupled with a machine learning approach, can provide many insights.

References

[1] Bureau of Labor Statistics, U.S. Department of Labor. National Longitudinal Survey of Youth 1979 cohort, 1979-2016 (rounds 1-27). Produced and distributed by the Center for Human Resource Research (CHRR), The Ohio State University. Columbus, OH: 2019.

[2] "Global Early Adolescent Study," 2021. https://www.geastudy.org/

[3] Hertel, Bradley and M. Hughes. (1987). "Religious Affiliation, Attendance, and Support for 'Pro-Family' Issues in the United States." Social Forces. 65: 858-882.

[4] Peek, Charles W., G. D. Lowe, and L. S. Williams. (1991). "Gender and God's Word: Another Look at Religious Fundamentalism and Sexism." Social Forces 69 (4): 1205–1221.

[5] Suftin, Erin L., M. Fulcher, R. P. Bowles, and C. J. Patterson. (2008). "How Lesbian and Heterosexual Parents Convey Attitudes about Gender to Their Children: The Role of Gendered Environments." Sex Roles 58 (7-8): 501-13.

[6] Lukes, Dylan. (2018). "Revisiting the Long-Term Impacts of Head Start: A Machine Learning Approach."

[7] Davis, Shannon and T.N. Greenstein. (2009). "Gender Ideology: Components, Predictors, and Consequences." Annual Review of Sociology 35: 87-105.

[8] Brand, J.E., B. Koch, P. Geraldo, and J. Xu. (2021). "Uncovering Sociological Effect Heterogeneity Using Tree-Based Machine Learning." Sociological Methodology 51 (2): 189-223.

[9] Beck, Amir, and M. Teboulle. (2009). "A Fast Iterative Shrinkage-Thresholding Algorithm for Linear Inverse Problems." SIAM Journal on Imaging Sciences 2 (1): 183-202.

[10] Janjusevic, Nikola. "Understanding ISTA as a Fixed-Point Iteration," February 29, 2020. https://nikopj.github.io/blog/understanding-ista/

[11] Gramfort, Alexandre. "Lasso with ISTA and FISTA," 2015. https://gist.github.com/agramfort/ac52a57dc6551138e89b

A Appendix

Complete List of Features by Category

- What circumstances did the child grow up in?
  - Did child's father live in the household (before and at time of Y)
  - Total number of members in the household (before and at time of Y)
  - Poverty status of the household the child grew up in (before and at time of Y)
  - Number of siblings the child has (at time of Y)
- What are the gender attitudes of the child's mother?
  - Mother's responses to 5 questions capturing gender attitudes measured over time (at time of child's birth and at time of Y)
- What might influence the gender attitudes of the mother?
  - Race
  - Gender
  - Citizenship status
  - Mother's age at birth of child
  - Religion in which mother was raised (captured twice over survey period)
  - Current religious affiliation (captured twice over survey period)
  - Frequency of religious service attendance (before and at time of Y)
  - Employment hours (at time of Y)
  - Highest grade completed
  - Number of jobs held (at time of Y)
- Characteristics of the child
  - Sex
  - Race
  - Age at survey (at time of Y)
  - Child's self perception/self-worth (captured repeatedly over survey period)

Figure 1

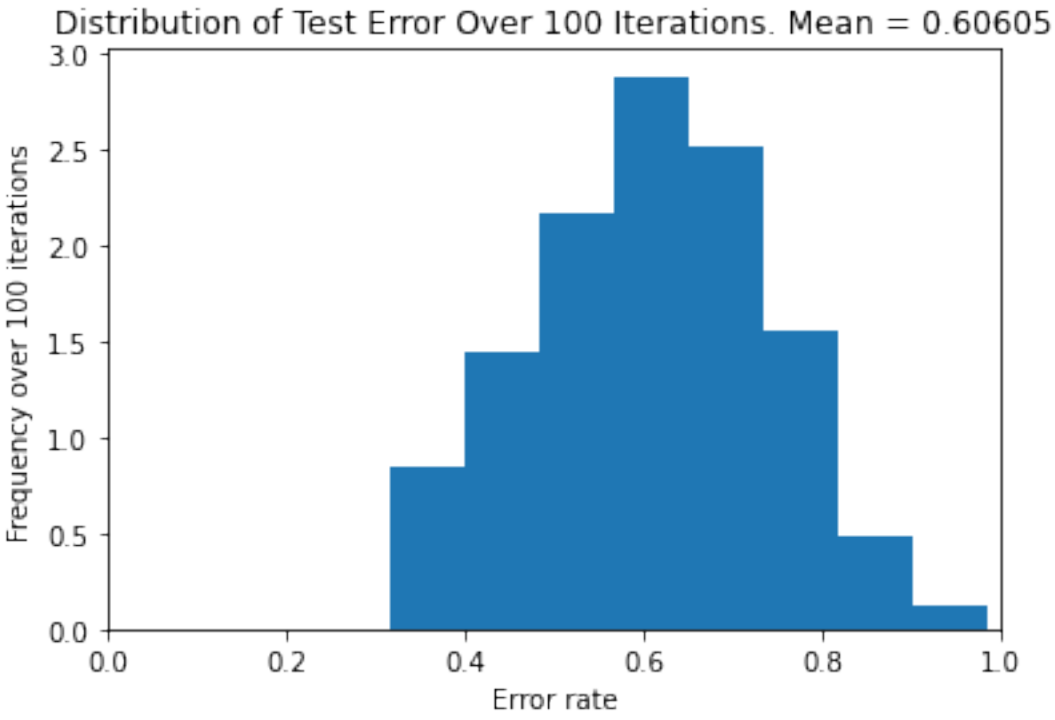


Figure 2

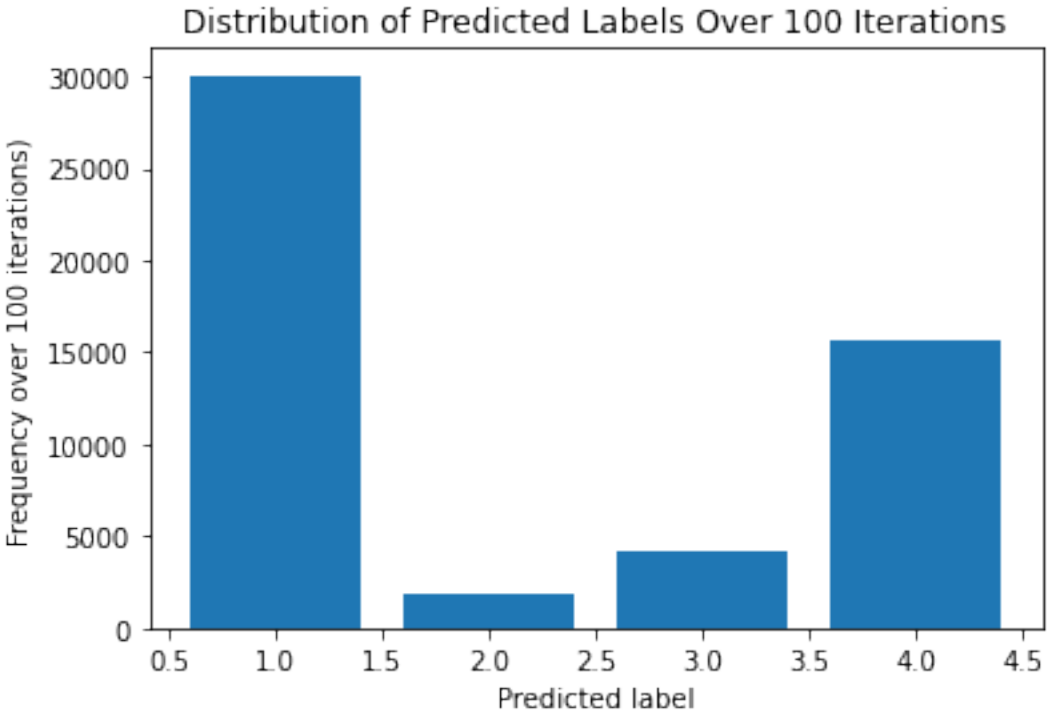


Figure 3

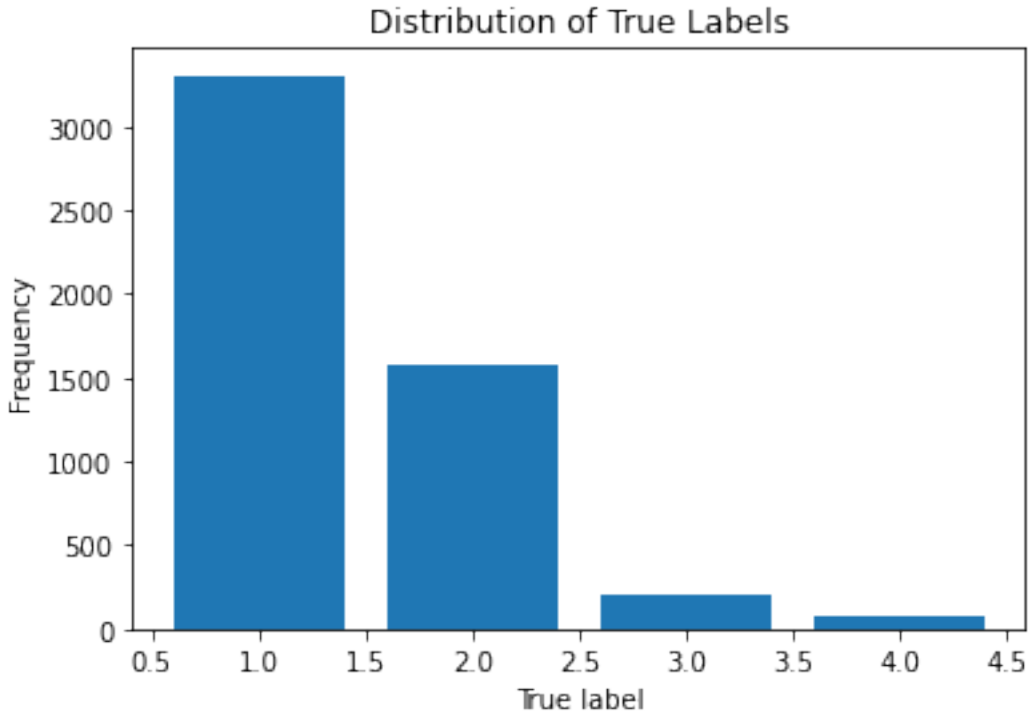


Figure 4

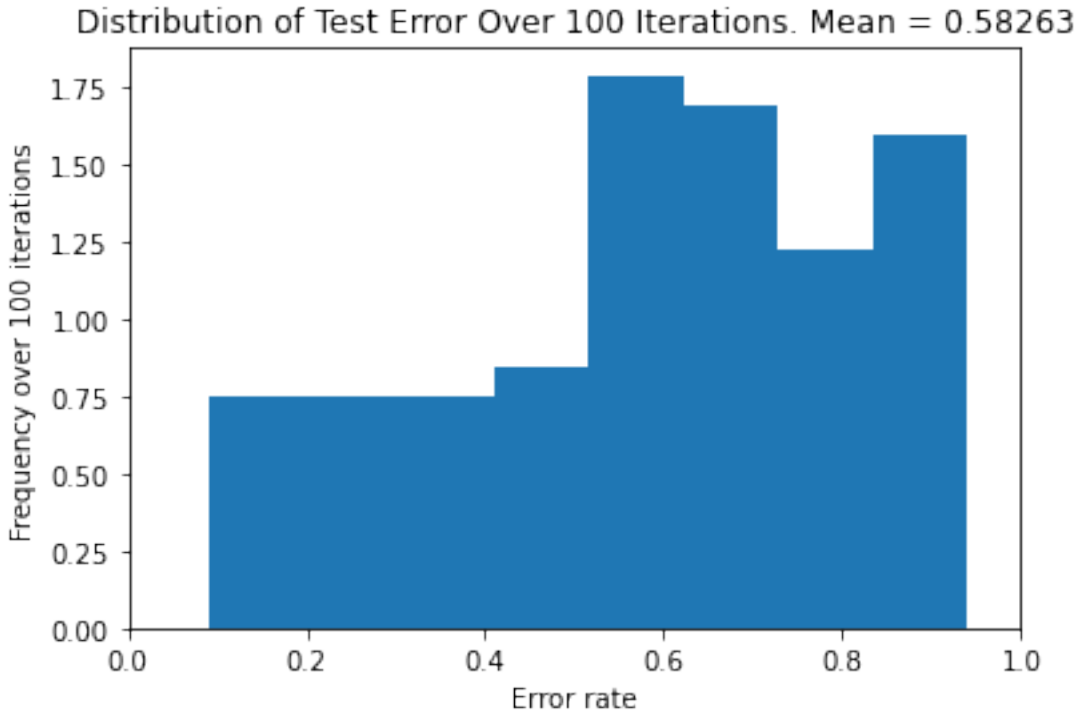


Figure 5

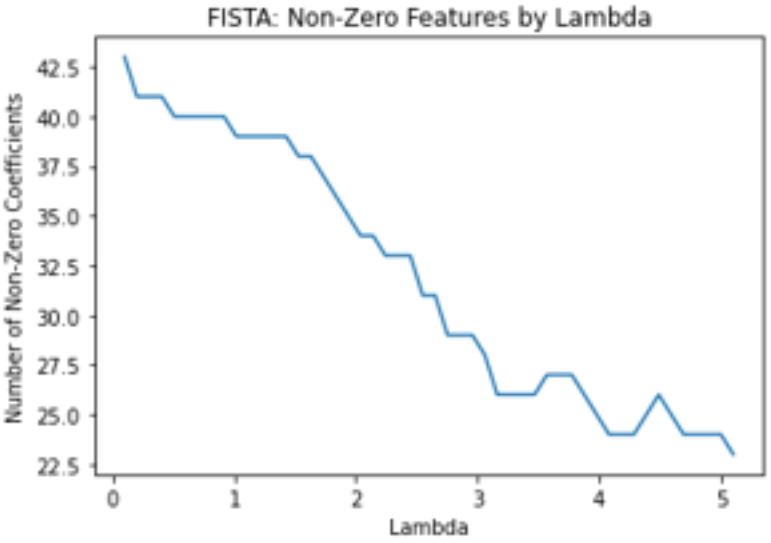


Figure 6

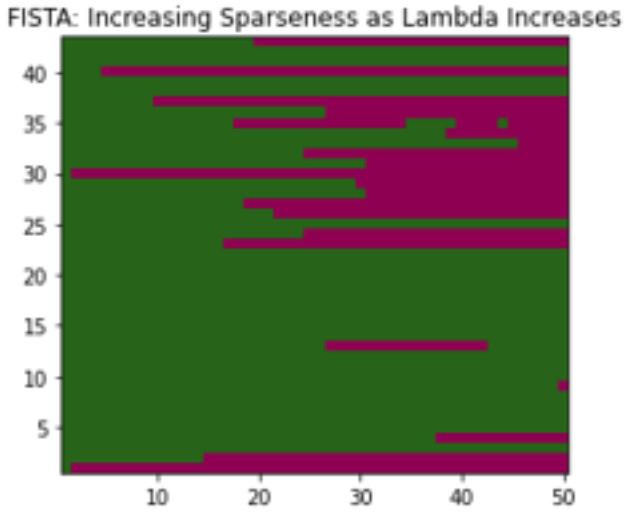
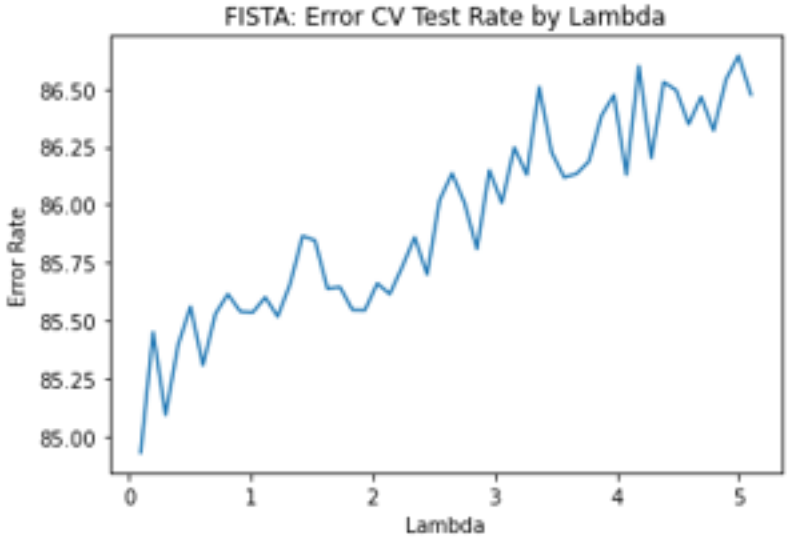


Figure 7



Lasso Code Snippet

```
# code framework draws guidance from
# Gramfort, Alexandre. "Lasso with ISTA and FISTA," 2015. https://gist.github.com/agramfort/ac52a57dc6551138e89

def soft_thresh(x, l):
    '''
    Implement soft-thresholding.
    args:
    - x: weights
    - l: regularization parameter

    Returns output from proximal function.
    '''
    return np.sign(x) * np.maximum(np.abs(x) - l, 0.)


def ista(A, b, l, maxit):
    '''
    Implement ISTA algorithm for finding Lasso solution.
    args:
    - A: data
    - b: variable to predict
    - l: regularization parameter
    - maxit: maximum number of iterations

    Returns: vector of weights
    '''
    x = np.zeros((A.shape[1], 1)) # initial guess
    L = linalg.norm(A) ** 2  # Lipschitz constant
    for _ in range(maxit):
        # sub-gradient in proximal function
        x = soft_thresh(x + np.dot(A.T, b - A.dot(x)) / L, l / L) # prox(sub-gradient)
    return x

def fista(A, b, l, maxit):
    '''
    Implement FISTA algorithm for finding Lasso solution.
    args:
    - A: data
    - b: variable to predict
    - l: regularization parameter
    - maxit: maximum number of iterations

    Returns: vector of weights
    '''
    x = np.zeros((A.shape[1], 1)) # initial guess
    t = 1 # scalar
    z = x.copy()
    L = linalg.norm(A) ** 2  # Lipschitz constant
    for _ in range(maxit):
        xold = x.copy() # x from previous iteration introduces accelaration
        z = z + A.T.dot(b - A.dot(z)) / L # same as ISTA algorithm
        x = soft_thresh(z, l / L) # z depends on x_old
        t0 = t
        t = (1. + sqrt(1. + 4. * t ** 2)) / 2.
        z = x + ((t0 - 1.) / t) * (x - xold) # here, x is a function of previous x (not seen in ISTA)
    return x
```