Addressing Inequity Through Modeling:

Updating Public Defense Funding Models In Washington State

Abstract

The Washington State Office of Public Defense (OPD) has identified multiple shortcomings in their county funding distribution methodology including: missing variables, disequitable funding allocation, difficulties in interpretability, and arbitrary or unfounded model coefficients. We consider and evaluate a least absolute shrinkage and selection operator (LASSO) model to attempt to address the considerations raised by the OPD. We, further, use unsupervised Principal Component Analysis (PCA) to reduce dimensionality and support our model creation. The goal of this report is to create an updated model for the OPD that addresses inequity in funding distribution through the inclusion of additional socioeconomic variables while simultaneously reducing model arbitrariness and increasing transparency.

Key words: Public Defense, Principal Component Analysis (PCA), Least Absolute Shrinkage and Selection Operator (LASSO), Budget Equity.

Background and Significance

Universal access to legal representation for indigent citizens is enshrined in the 6th amendment of the U.S. Constitution and stands as a core tenet of the judicial system. How to properly fund this unalienable right, however, has been a question the U.S. has struggled to answer since 1893, when Clara Shortridge Foltz first envisioned creating "a defense office for the public as a counterbalance to the office of the public prosecutor."[13] This approach, spearheaded by Foltz, finally mandated access to public defense, though the burden of funding was placed solely upon county governments. County-led funding remained preeminent until the Supreme Court's rulings in *Gideon v. Wainwright*[5] and *Argersinger v. Hamlin*,[1] wherein a new interpretation of the Sixth Amendment required state government support in public defense funding. Despite Supreme Court rulings, many states, including Washington, continued to operate largely decentralized and disunified systems of public defense.

Within Washington state, the decentralized public defense system creates difficulties in assessing how to best support public defense attorneys and resources.[4] Since every county has their own office(s) and there is no state-wide reporting system, it is impossible for the state to use an internal knowledge system when doling out its annual public defense budget to the counties. To counteract this problem, the Washington State Legislature passed RCW 2.70.005 in 1996 to create the Office of Public Defense (OPD), and asked all counties and interested municipalities to submit an application for state funding. The OPD funding application collects information on county office's caseloads, resources utilized, and on relevant county metrics, like population. The data the OPD collects informs the 10.101.060 model—so named because the model was created through passing RCW 10.101.060.[8] It is also important to note that while a decentralized system enables counties to express their unique priorities through funding allocation, such a system allows for inequity to consequently arise.

The 10.101.060 funding model, created in 2005 and actively used to allocate the public defense budget in Washington State, only considers county *misdemeanor caseload* and *population* when determining distribution—prior to 2005, only population was considered.[9] The bivariate consideration is then used to inform 10.101.060 model as follows: 47% of total funds are disbursed based on proportional county population, 47% is determined by misdemeanor caseload numbers, and the remaining 6% is divided evenly among counties to ensure no county goes unfunded (**Figure 1)**.[8] This process is relatively unsophisticated and unfounded from a statistical modeling perspective, and clearly leaves room to increase interpretability and relevance. OPD employees, for instance, are unclear as to why one thirty ninth of 6% is the baseline funding amount by county. More pressingly, however, under the 10.101.060 model, counties may receive the same amount of funding regardless of factors which are known to affect people's access to legal representation, like income and poverty rates.[12] **Figure 2**, for illustration of this point, shows how OPD funds are not equitably allocated to counties with higher rates of poverty.

Research Question

Given the shortcomings of the 10.101.060 model, and the legal and societal importance of its mandate, we hope to revise the OPD funding model to: (1) rely on statistical intuition and reasoning in place of arbitrary metrics, (2) explain how the values are obtained to increase transparent in application, and (3) account for additional variables such as poverty, median income, and average home prices. In short, we ask: What is the optimal way to disburse state funds to county public defenders offices in Washington?

Methods

To answer our question, we rely on data from OPD and the [University of Washington] about caseload resources and county statistics from 2018 to 2020.[2] We wrangled the collected data from display to analytic form, refactoring and mutating variables when appropriate to increase utility and to eliminate unhelpful categorical variables. While we hoped to construct models starting from 2005 (when the 10.101.060 model began), high-quality data—without missing responses for some or all counties—is only present from 2018 to 2020. Thus, we trained our model on data from 2018 with all 21 predictors listed on **Table 2**, and tested it on data from 2019, thereby predicting 2020 budget proportions. These proportional predictions are useful, given how we hope to optimally distribute state funds from a preexisting budget that is subject to change each year. The variables we used can be partitioned into three categories: income/earnings, juvenile crime, and adult crime by county.

When selecting a model type, we primarily wanted a model which would include more than misdemeanor caseloads and population as variables. Secondarily, we wanted a model that would perform feature selection and address collinearity between certain variable groupings, while being cautious about the difficulty of defining a response variable. We ultimately decided, given these desired parameters, on a Least Absolute Shrinkage and Selection Operator (LASSO) model. While we compared model predictions to the *actual* budget allocations within LASSO, the motivating cause of our project is to redefine the model, which is why we also used PCA. As an unsupervised learning technique, PCA does not rely on a response variable, unlike LASSO. It is, therefore, a useful analysis to consider for our purposes, as we hope to create a completely distinct, new form of budget allocation. It is worth mentioning, however, that a drastic change in funding allocations will likely not be passed by the legislature, and thus a low rMSE that accounts for more variables, especially poverty, is crucial. To achieve this, we implemented both LASSO and PCA in R, using the `glmnet` and `recipes` packages. These two methods can reduce the dimensionality of a dataset, with the difference being that LASSO selects certain predictors and PCA includes all features.

As a brief overview, LASSO automatically selects significant predictors to generate a sparse linear regression model. It is used to estimate the least squares of linear regression with the addition of a shrinkage penalty.[6] PCA reduces the dimensionality of a dataset but minimizes information loss to keep dataset's variability.[6] A more in-depth mathematical exploration of LASSO and PCA can, respectively, be found in **Appendix 1.1** and **Appendix 1.2**.

Results

LASSO

In this section, we look at the specific output of our LASSO model. The model, notably, had nine non-zero predictors, listed in **Table 1**. This insight is particularly helpful for our purposes, as we see the variables for misdemeanor cases and population still serving as important predictors of the 2020 budget, though the model also considers several additional factors. This suggests that there are other uncorrelated variables that could enhance how Washington state allocates budgeting to its counties. These impact how the model makes predictions and, accordingly, supports the argument of including more variables in the OPD funding model. The best LASSO tuning parameter, as determined through cross-validation and calculating the minimum mean-squared error was λ = 0.0010476. A

Table 1: Non-Zero LASSO Model Coefficients

Variable	Coefficient
Intercept	-0.0286993
Population_2019	0.0000002
Percent_Individuals_below_poverty_level_2015_2019	0.0500727
Adult_felony_cases_per_1000	0.0006815
Misdemeanor_cases_filed_2019	-0.0000036
Misdemeanor_cases_per_1000	0.0005871
Juvenile_offender_cases_per_1000	0.0000894
Juvenile_offender_cases_assigned_to_counsel	0.0000128
Juvenile_Offender_Appointments	-0.0019388

visualization of the mean-squared error across the lambda grid is shown in **Figure 3.** For more information on how this λ value impacts the predictors, please reference **Figure 4**, which shows the shrinkage of coefficients across all lambda values contained in our grid.

When considering more variables, the LASSO model significantly outperforms the 10.101.060 model, as it more fairly distributes funds to counties with higher poverty levels. For instance, consider these pairs of counties: Okanogan and Pacific, and Kittitas and Klickitat. Both Okanogan and Kittitas have high rates of poverty, but received approximately the same proportion of the public defense budget as the other under the 10.101.060 model (**Figure 5**). The new predictions from LASSO, however, indicate more equity in their distributions, with Okanogan and Kittitas receiving more funds than their less impoverished counterparts.

Principal Component Analysis

PCA helps us achieve our goal by creating principal component vectors, which collapse multiple features into fewer dimensions. We can, therefore, create a model that includes these principal components, accounting for multiple variables, without worrying about increased dimensionality. As PCA includes all of our features when projecting new values, it has an advantage over LASSO, which selects away certain features. Armed with more information about each of the counties, we argue that models created using the principal components in place of the regular features will encompass more variables and thus, make more equitable budget allocations. As **Figure 6** shows, the second principal component suggests a numerical relationship by county that is informed by more than just misdemeanor caseloads or population. The first principal component generally follows population by county, though not perfectly—such as Snohomish scoring higher than Pierce. This suggests that budget allocation predictions made by a model using PCA features will better address poverty, median income, median home prices, and other resources available to public defenders in Washington State. Without relying on the crutch of the preexisting proportions of the budget distributed to each county, which the LASSO model uses to determine which variables should be included and the shrinkage penalty, PCA allows for an unsupervised learning technique that focuses exclusively on utilizing all available data to make budget allocation decisions.

Discussion and Conclusions

The model and analysis we present here was created and trained on limited data, meaning it would be inappropriate to claim that we have built the best possible model for disbursing OPD funds to counties in Washington. We can, however, assert that we created an improved, statistically grounded model with the available resources that successfully increased model transparency and interpretability, and accounted for a greater variety of socioeconomic factors. PCA, further, reduced the dimensionality and number of features present in the data, allowing for us to consider all features without the same high concern for collinearity or feature selection excluding a necessary variable. While other machine learning techniques—such as boosted trees and LASSO—rely on excluding various predictors, a budgeting model necessitates consideration of as much information as possible when making allocations, meaning PCA has strong future potential. A clear avenue to improve our work, naturally, is to collect and analyze additional data from a wider range of sources and years. Another more difficult, but potentially transformative update to our work would be to secure outside funding distribution models with which to compare to ours. We hope that our work can be used—in some small part—to help update and modernize the criminal justice system to best support the needs of counties in a more equitable, transparent manner.

References

1. “Argersinger v. Hamlin” (1972), Supreme Court.
2. Carney, K., and Morales, A. (2020), “Washington counties by population,” Washington Demographics, United States Census Bureau.
3. Coplen, D. (2012a), “Washington state office of public defense,” Washington State Office of Public Defense.
4. Coplen, D. (2012b), “Public defense improvement program,” Washington State Office of Public Defense.
5. “Gideon v. Wainwright” (1963), Supreme Court.
6. Huang, S., and Houtao, D.(2021), Data Analytics: A Small Data Approach, Chapman and Hall/CRC.
7. James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013), An introduction to statistical learning, Springer.
8. Legislature, 59th (2005a), “RCW 10.101.060 Appropriated funds-Use requirements.” https://app.leg.wa.gov/RCW/default.aspx?cite=10.101.060.
9. Moore, J. I. (2018), Annual fiscal report, Washington State Office of Public Defense.
10. Ogletree, C. J. (1995), “An essay on the new public defender for the 21st century,” Law and Contemporary Problems, JSTOR, 58, 81–93.
11. Legislature, 59th (2005b), RCW 10.101.070: County monies, Washington State Legislature.
12. Rhode, Deborah L. (1992), Access to Justice. Oxford University Press.
13. “Understanding Gideon's impact, part 2: The birth of the public defender movement” (2021), Sixth Amendment Center, Inc.

Appendices

Figure 1 – Graphical Illustration of the 10.101.060 Funding Model

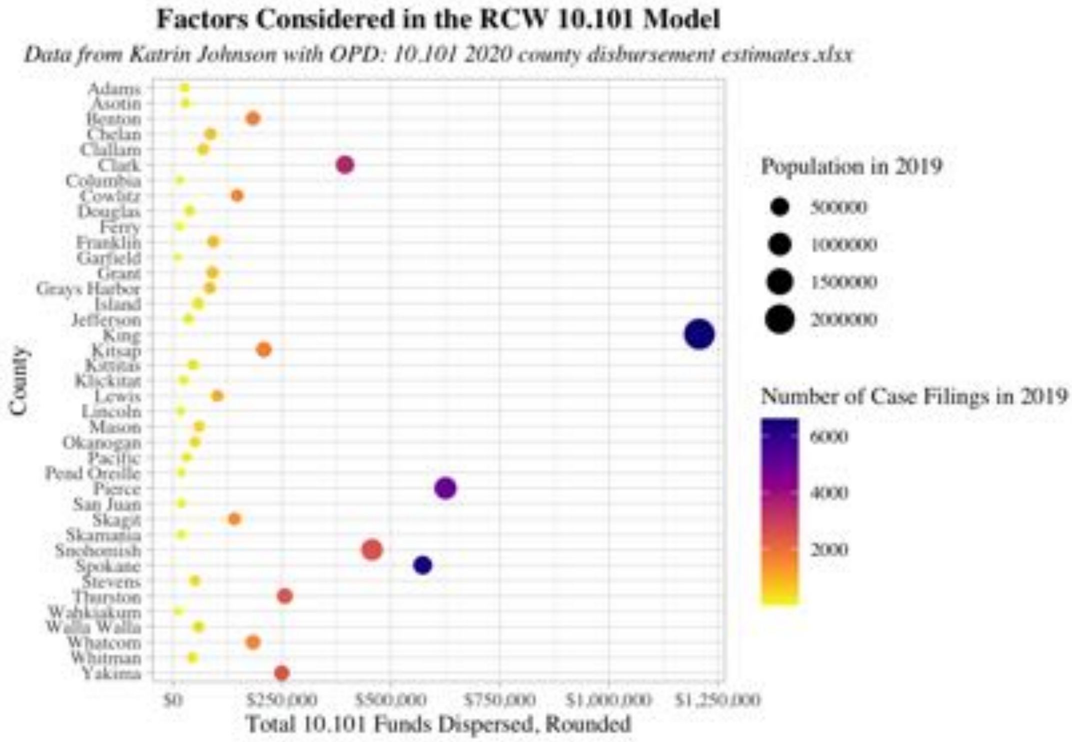

Figure 2 – Assessing 10.101.060 Funding Distribution by Poverty Rates

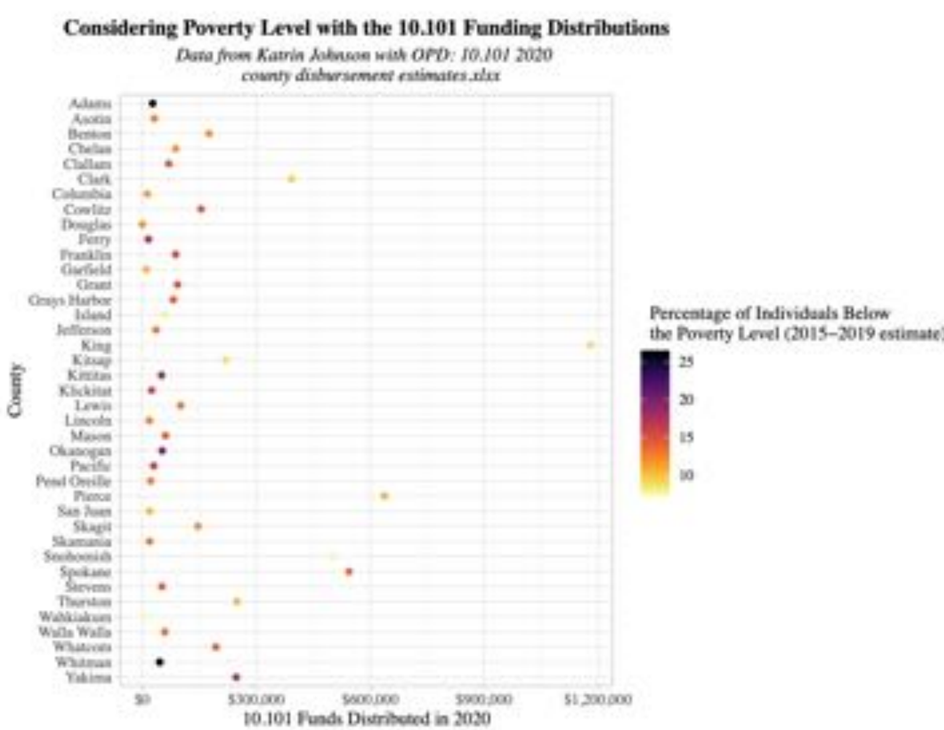

Figure 3 – Mean-Squared Error by Shrinkage Penalty for LASSO

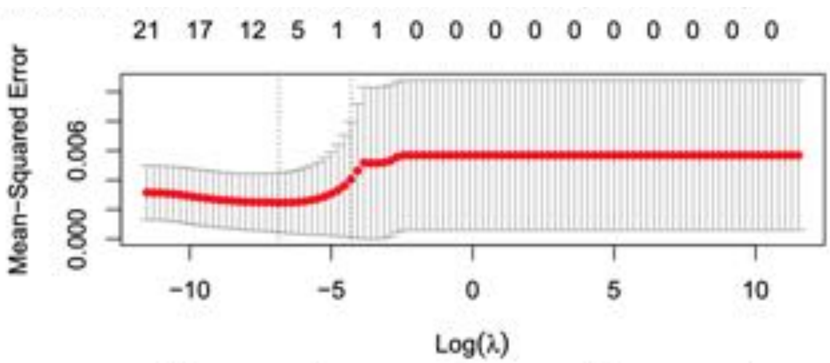

Figure 4 – Shrinkage Penalty for Predictors Included in LASSO

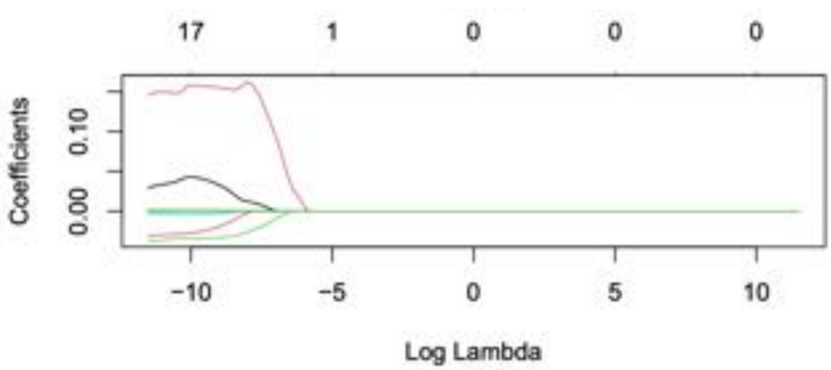

Figure 5 – Comparing Funding Allocation by Rates of Poverty

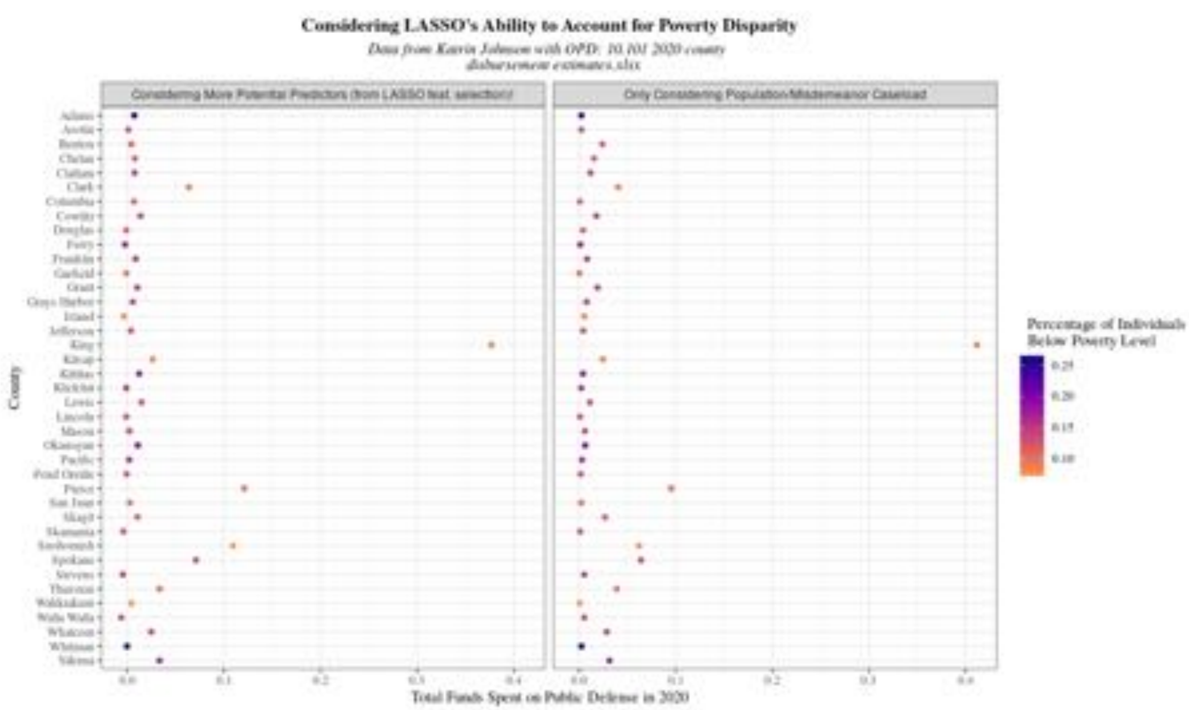

Figure 6 – PCA Plot

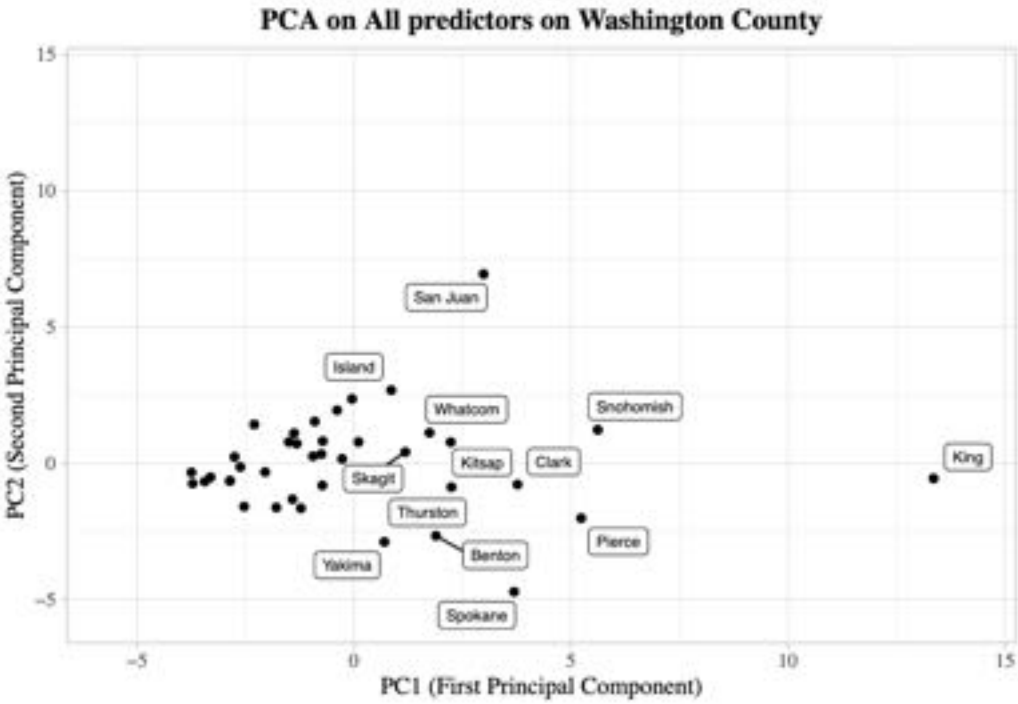

Table 2 – All Predictors Used for Modeling

Population_2019
Percent_Individuals_below_poverty_level_2015_2019
Median_household_income_2015_2019
Adult_felony_cases_filed_2019
Adult_felony_cases_per_1000
Adult_felongy_cases_assigned_to_counsel
Misdemeanor_cases_filed_2019
Misdemeanor_cases_assigned_to_counsel
Misdemeanor_cases_per_1000
Juvenile_offender_cases_filed_2019
Juvenile_offender_cases_per_1000
Juvenile_offender_cases_assigned_to_counsel
Adult_Felony_Appointment_Rates
Adult_Misdemeanor_Appointment_Rates
Juvenile_Offender_Appointments
Median_Home_Price_Q4_2018
Median_Home_Price_Q1_2019
Median_Home_Price_Q2_2019
Median_Home_Price_Q3_2019
Median_Home_Price_Q4_2019
Median_Home_Price_Percent_Change_by_Year

Appendix 1.1 – LASSO model

The formula for LASSO is used to estimates the least squares ($\hat{\beta}$) of linear regression with the addition of penalty values. The formula for estimating the least squares of linear regression is $\hat{\beta} = arg\,min(y - X\beta)^T(y - X\beta)$. Where $y \in \mathbb{R}^{N \times 1}$ is the measurement vector of the outcome, $X \in \mathbb{R}^{N \times p}$ is the data matrix of measurement vectors (N) of predictors (p), and the regression coefficient vector is $\beta \in \mathbb{R}^{p \times 1}$. Thus the formula for LASSO is:

$$\hat{\beta} = arg\,min\{(y - X\beta)^T(y - X\beta) + \lambda||\beta||_1\}$$

where $||\beta||_1 = \sum_{i=1}^{p} |\beta_i|$, and norm penalty has penalty parameter (λ).[6]

Appendix – 1.2 PCA

PCA is a multi-step process to obtain two principal components. It starts by computing the d dimensional dataset to get the covariance matrix. This will allow the calculation of eigenvectors and its corresponding eigenvalues. One then chooses the k eigenvectors with the largest eigenvalues from the calculation to generate a $d \times k$ dimensional matrix W. This matrix will be used to surjectively transform the samples to the new subspace.[6] Finally, the process will generate two principal components and project the data points onto the new subspace.[6]