

DATA 558: Homework 3

Trevor Nims

April 18, 2021

Exercise 1

Given that we estimate the regression coefficients in a logistic regression model by minimizing:

$$F(\beta) = \frac{1}{n} \sum_{i=1}^n \log(1 + \exp(-y_i x_i^T \beta)) + \lambda \|\beta\|_2^2$$

for a particular value of λ in the interval $(0, \lambda_{\max}]$, with $\lambda_{\max} \gg 0$, I will describe some interesting behavior of the model across the range of our λ values.

(a)

As λ decreases from λ_{\max} to 0, the misclassification error on the training set will steadily decrease (ii). Initially, with $\lambda = \lambda_{\max}$ we will have a very high penalty on large coefficients, causing our optimal $\beta = \hat{\beta}$ to be close to the zero vector. This will create a highly biased and unpredictable model with low flexibility and therefore low accuracy. As we decrease the size of λ and it approaches 0, however, our model begins to act more and more like one based on an unregulated objective function. By not imposing any penalty on the coefficient size, when $\lambda = 0$ we will have a model that will produce the highest training set accuracy.

(b.)

As λ decreases from λ_{\max} to 0, the misclassification error on the test set will decrease initially, and then eventually start increasing in a U shape (iv). Initially, the case is the same as it was with the misclassification error on the training set. With $\lambda = \lambda_{\max}$ we will have a very high penalty on large coefficients, causing our optimal $\beta = \hat{\beta}$ to be close to the zero vector. This will create a highly biased and unpredictable model with low flexibility and therefore low accuracy. However, accuracy on the test set will only steadily increase to an optimal $\lambda = \hat{\lambda}$ with $0 < \hat{\lambda} < \lambda_{\max}$. After that optimal $\hat{\lambda}$ is reached, the test set accuracy will begin to decrease at a rate proportional to $\|\beta\|_2^2$ as the model becomes under-penalized and begins to overfit to the training data.