

558 Homework 3 - Exercise 1

Anushna Prakash

April 23, 2021

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

What happens to the misclassification rate when $\lambda \gg 0$?

- (i) Steadily increase.
- (ii) Steadily decrease.
- (iii) Increase initially, and then eventually start decreasing in an inverted U shape.
- (iv) Decrease initially, and then eventually start increasing in a U shape.
- (v) Remain constant.

(a) When λ decreases from λ_{max} to 0, the misclassification error on the training set will:

- ii. Steadily decrease.

We can think of the objective function as being composed of a goodness-of-fit measure, which is contained in the logit function, and a model complexity penalty in the $\lambda \|\beta\|_2^2$ term. With very high λ , the model is being punished heavily for the size of its coefficients, and so the coefficients will be pulled closer to 0. This can combat overfitting on the goodness-of-fit measure, which is what minimizes the misclassification. As the penalty decreases, the model is allowed to maximize goodness-of-fit on the training set, and so the misclassification rate will decrease.

(b) Suppose now that we compute the misclassification error on a large test set consisting of previously-unseen observations drawn from the same distribution as the training set. Which of the patterns (i) – (v) do we expect to observe as λ decreases from λ_{max} to 0?

- iv. Decrease initially, and then eventually start increasing in a U shape.

A λ that is too high will heavily penalize the coefficients in the model and bring them closer towards 0. The model will be highly biased towards this overly simplistic model with very small coefficients, and will consequently have a high misclassification rate. However, as the penalization decreases, goodness-of-fit is prioritized slightly more, and so the misclassification rate on a test set will decrease. However, there is tradeoff to allowing more variance in the coefficients into the model and decreasing the bias, and after a certain point the model will have overfitted on the training data and will not be generalized well enough to perform on new data. This is the point at which decreasing λ will have the effect of raising the misclassification rate again, giving us a U-shaped curve.