

MATH 392 Problem Set 9

1. GLM with Gaussian Response

Consider the special case of the generalized (simple) linear model where we assume independent Gaussian errors and are linked to the linear predictor using the identity function (vanilla simple linear regression). We can summarize that model as follows (note that in the notation used in this problem, everything is a scalar):

- i. If $X = x$, then $Y = \beta_0 + \beta_1 x + \epsilon$ where the β_j are (unknown) parameters and the ϵ is a random variable.
- ii. $\epsilon \stackrel{iid}{\sim} N(0, \sigma^2)$ for some (unknown) parameter σ^2 .

When this model was first derived, and only (i) was assumed, the parameters were estimated by minimizing the residual sum of squares (these are the least-squares estimates, $\hat{\beta}^{LS}$). This yielded

$$\begin{aligned}\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ \hat{\beta}_1 &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2} \\ \hat{\sigma}^2 &= \frac{1}{n-2} \sum_{i=1}^n (y_i - (\hat{\beta}_0 + \hat{\beta}_1 x))^2\end{aligned}$$

Now that we have added (ii), we have specified a full density function for the random variable Y , $f(y|X = x, \beta_0, \beta_1, \sigma^2)$, which enables a familiar route to estimation: maximum likelihood.

Find the maximum likelihood estimates of β_0 , β_1 , and σ^2 .

- a) Provide the derivation of closed-form solutions, if they exist. For reference, if $X \sim N(\mu, \sigma^2)$, $f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$.
- b) Describe in pseudocode (or actual R code) how to find them using numerical optimization.

2. Logistic Regression MLEs: Bias, Variance, and Shrinkage:

For this problem you'll be working in a setting when the design matrix including the intercept is an $n \times 2$ matrix X and the response is Bernoulli with an inverse logit link function to the linear predictor (logistic regression). Since there is no closed form of the MLE, you'll be using simulation, meaning you'll need to specify values for all of the parameters needed to generate data from the Logistic Regression model.

- a) *How does a single estimate compare with the true mean function?* Simulate one data set and fit one model using the MLEs. Construct a scatterplot with the simulated data, the estimated mean function ($\hat{E}(Y|X = x)$) and the true mean function ($E(Y|X = x)$).
- b) *Is the MLE Biased?* Simulate many data sets and fit many models using MLE. Create a plot similar to the previous, but with *all* of the fitted models' mean functions plotted. To make this more complex plot intelligible, I recommend the `gghighlight` package.
- c) *How does the bias of an estimate change with sample size for a particular value of the parameter?* For a single fixed value of β_1 , construct a plot that shows the relationship between n and the bias of the corresponding MLE.
- d) *How does the bias of an estimate change with sample size for multiple values of the parameter?* Extend the idea of the previous plot by expressing the relationship between the value of β_1 and the corresponding element of $\hat{\beta}^{MLE}$ for various fixed values of β_1 . Examine this relationship at a handful of sample sizes n .

- e) *Can I perform shrinkage on logistic coefficients?* The original motivation for ridge regression was to make the $X'X$ matrix in OLS invertible. Statisticians have since realized the practical value of its variance-reducing characteristics when shrinking $\hat{\beta}$ towards zero. Traditional ridge regression is performed by adding a penalty term, $\lambda \sum_{j=1}^p \beta_j^2$, to the RSS. In logistic regression, we instead of finding our estimates by minimizing RSS, we choose to maximize the likelihood. Although the original motivation of matrix inversion is lost, it can still be perfectly valid and valuable to shrink the logistic regression estimates by penalizing the likelihood.

For the same data set that you used to create the plot in part a, find three more estimated mean functions, each one corresponding to a different value of λ , and add them to the plot. Admittedly, this will demonstrate the *downside* of penalized regression: that we have actually *increased* the bias. The plot should have five overlaid sigmoid curves. Play around with the values of λ so that you can see the shape of all five on the same plot.

- f) To bring in a sense of both bias and variance, select one of your values of λ and use it to replicate the plot from part b, but now with two clumps of sigmoids: one corresponding to the MLE, the other to the ridge estimates. Use color to differentiate between the two clumps. Describe what the plot demonstrates about the bias and variance for the MLE and the ridge estimates in this setting.