4.2: Variations on OLS (Ordinary Least Squares)

Dr. Bean - Stat 5100

If Robust Regression is less sensitive to outliers and influential points, why should we *ever* use ordinary least squares (OLS) regression and worry about outliers and influential points?

IF assumptions regarding residuals are satisfied, then OLS parameter estimates truly are the "best" (i.e. unbiased with minimum variance). Robust regression involves the (somewhat arbitrary) selection of a weighting function, which affects the estimates and the statistical inference of those estimates. Robust regression is also computationally more expensive. In short, if a simpler solution is equally effective in solving the problem, we should always prefer the simpler solution.

Metaphor: You can technically use a sledgehammer to pound in a nail, but it will likely be a little messy. Just because a sledgehammer can "do the job" of a regular hammer does not diminish the value of owning a hammer.

In what ways is non-linear regression different from simply adding higher ordered predictors (i.e. X^2 or X^3) in our model?

Higher ordered predictors are still linear for the beta coefficients. Thus higher ordered terms can only describe a limited number of non-linear relationships. We use non-linear regression when the domain theory suggests that the model form should be non-linear in the betas.

Why would we want to avoid "making up" or own non-linear model forms for non-linear regression? Why depend so much on the domain theory for the model form?

We want the non-linear model to have physical meaning. While our predictions from the made up model might look OK, the estimated parameters will likely have no meaningful interpretation.

What are some potential issues with the use of gradient descent to find optimal model coefficients?

- speed: It can take a long time to reach the optimal solution when you are taking small steps at each iteration.
- local extremes: The gradient descent cannot recognize if the solution it converged to is the global minimizer of the loss function, or just a *local* minimizer.

The residuals of your OLS regression model show signs of heteroskedasticity. You are able to identify a transformation for Y that fixes the issue. Should you use OLS with the transformed Y or use weighted least squares regression? Please give arguments in favor of both methods.

- Argument for OLS: simpler model form, ability to conduct traditional statistical inference. Easier to compute.
- Argument for weighted least squares: avoid the use of a transformation, which makes model coefficients easier to interpret.