High-Dimension Considerations

Wednesday, September 22, 2021 6:09 PM

- Regression in High Dimensions
  - If $p > n$ → Infinite # of solutions → Variance
  - If $p \approx n$ → Flexibility in Linear Models which can lead to overfitting

- As $p \to \infty$, and each feature is included in regression, $R^2 \to 1$, training $MSE \to 0$ regardless of whether features are related to response!

curse of dimensionality!

* In very high dimensional settings → $\hat{\sigma}^2$ estimates are not accurate
  - ↳ AIC, BIC, Adj R^2 are not accurate!
  - ↳ Ridge & Lasso solutions are constrained → one of many possible solutions!