

The Bias Variance Tradeoff and Double Descent

Background

The bias-variance tradeoff is a major theme in statistical learning. Essentially, this tradeoff is about the sources of error when testing a model. These can be shown to undersood in terms of three sources.

1. how sensitive the model is to variation in the training data
2. how complex the model is – models with fewer parameters are more biased and models with more parameters are less biased (more flexible)
3. Irreducible error which no one can do anything about

In the past 5-10 years, the rise of neural networks with very many parameters has called this tradeoff into question. This is because of a phenomena termed “double descent”. Basically, these “overparameterized” models perform better than expected given the bias-variance tradeoff. This has led to claims that the theory of machine learning needs to be rethought completely.

This lesson examines the issue primarily using the 2021 book *An Introduction to Statistical Learning with R* by Gareth James, Daniela Witten, Trevor Hastie and Robert Tibshirani. There is also a version of the book [with Python instead](#) by Gareth James, Daniela Witten, Trevor Hastie, Rob Tibshirani and Jonathan Taylor. We look at sections 2.2 and 10.8 in the R book.

Outline

1. 21. How Do We Estimate f ? Parametric and non-parametric models.
2. 24-25. The Trade-Off Between Prediction Accuracy and Model Interpretability
3. 29+. Assessing Model Accuracy
 - Mean Squared Error as a measure of quality of fit.
 - Train MSE vs Test MSE
4. 33. The Bias-Variance Trade-Off
 - Test MSE = Var + Bias + Error
 - Variance refers how easily the estimate changes as training data changes
 - Bias refers to error introduced by inherent expressive power of the model insofar as it matches the ground truth in the real world.
 - 36. “The challenge lies in finding a method for which both the variance and the squared bias are low. This trade-off is one of the most important recurring themes in this book.”
5. 439. Double Descent
 - Example with sine wave
 - 295. Piecewise Polynomials, Constraints and Splines
 - Interpolation threshold
 - Many fits to the data once past the threshold
 - see also [this interactive site](#) (disclaimer: The visuals are useful, I am not vouching for the text except insofar as it is illustrative of what many think.)
 - 442. Scope of Double Descent
 - Examine Figure 10.21 beyond the [-4,4]