LECTURE 6: EFFECTIVE DEGREES OF FREEDOM

REFERENCES

Lecture notes, handouts + Shalizi Sec 1.5.3.2 and Sec 3.4.3.

OVERVIEW

The notion of *effective degrees of freedom* gives us a quantitative measure of the complexity of an estimator, and a means of comparing error curves for different methods. Degrees of freedom is also directly related to the difference between expected test and training errors, the so-called *optimism* of the estimator; hence, by estimating the degrees of freedom we can construct estimates of expected test error as faster alternatives to cross-validation.

QUESTIONS

1. How does one define the degrees of freedom of a regression model? What assumptions are we making here?
2. What are the degrees of freedom of a linear smoother? How is this useful in practice?
3. How do you estimate the degrees of freedom if they cannot be calculated analytically?
4. What is the difference between in-sample and out-of-sample prediction errors?

DEFINITIONS AND NOTATIONS

1. Degrees of freedom
2. Linear smoother
3. Residual bootstrap
4. Optimism and Mallow's C_p statistic
5. Generalized cross-validation.

TOPICS

1. Degrees of freedom: Motivation, definition and examples
2. Estimating degrees of freedom via the bootstrap
3. (Part II) Using degrees of freedom to estimate the prediction error

WHAT'S NEXT?

Lecture 7: More on Kernel Regression (Shalizi Chapter 4 and Sections 1.5 and 9.3)