Lecture 4: MSE, Prediction Error Github Tools

Big Data and Machine Learning for Applied Economics Econ 4676

Ignacio Sarmiento-Barbieri

Universidad de los Andes

August 20, 2020

Agenda

Recap

- We started shifting paradigms
- Linear Regression
- Prediction vs Estimation
- Train and Test Samples
- Example in R

Motivation

- Working model is

$$y = f(X) + u \tag{1}$$

- Linear regression is the "work horse" of econometrics and (supervised) machine learning.

$$y = X\beta + u \tag{2}$$

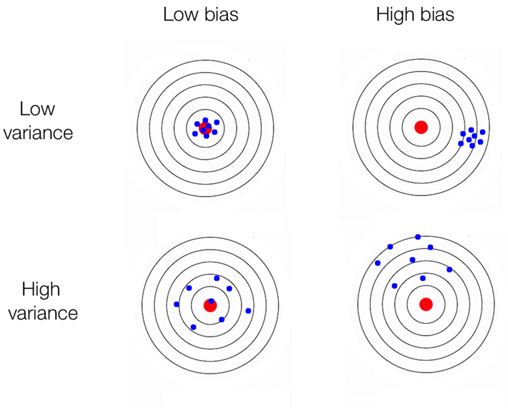
  - All the interest is on β
  - Gauss-Markov Theorem says that under classical assumptions it is BLUE

Mean Square Error

$$
\begin{aligned}
MSE(\beta) &= E(\hat{\beta} - \beta)^2 && (3) \\
&= E(\beta - E(\hat{\beta}))^2 + Var(\hat{\beta}) && (4)
\end{aligned}
$$

- Intuitively, the result says that how wrong is the estimate (MSE) depends on:
  - how uncentered it is (bias) and
  - how dispersed it is around its center (variance).

Mean Square Error



Source: https://tinyurl.com/y3xlh87o

Prediction Error

- Now suppose that the goal is to predict Y with another random variable $\hat{Y}$.
- The *prediction error* is defined as:

$$Err(\hat{Y}) \equiv E\left(Y - \hat{Y}\right)^2 \tag{5}$$

- Conceptually the prediction error is equal to the MSE
  - MSE compares a RV ($\hat{\beta}$) with a parameter (β)
  - $Err(\hat{Y})$ involves two RV

Prediction Error

Then

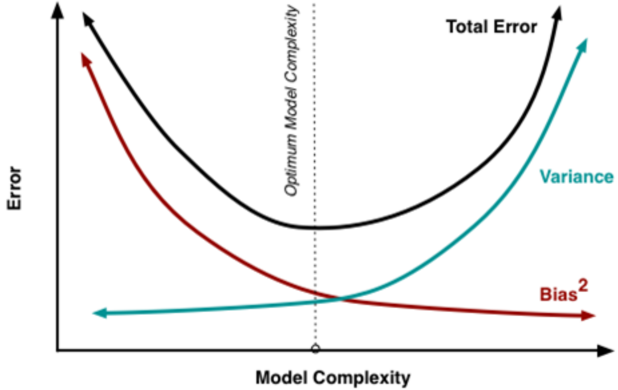
$$Err(\hat{Y}) = E(Y - \hat{f})^2 \quad (6)$$
$$= MSE(\hat{f}) + \sigma^2 \quad (7)$$
$$= Bias^2(\hat{f}) + V(\hat{f}) + \sigma^2 \quad (8)$$

Two parts:

- the error from estimating f with $\hat{f}$. (*reducible*)
- the error from not being able to observe u. (*irreducible*)

This is an important result, predicting Y properly we need a good estimate of f.

Prediction Error



Source: https://tinyurl.com/y4lvjxpc

A very interesting discussion in a recent Twitter thread by Daniela Witten:

https://twitter.com/daniela_witten/status/1292293102103748609?s=20

Prediction Error

- Model $y = X\beta + u$
- $\hat{y} = \hat{X}\beta$ is the prediction
- The estimated prediction error is

$$\hat{E}rr(\hat{Y}) = \sum(y_i - \hat{y}_i)^2 \tag{9}$$

- Common alternatives involve: the mean or the square root
- In Econometrics

$$\hat{E}rr(\hat{Y}) = \sum_{i=1}^{n} e_i^2 \tag{10}$$

- $R^2 = 1 - \frac{Err(\hat{Y})}{TSS}$
- OLS minimizes $\hat{E}rr(\hat{Y})$ and maximizes $R^2 \rightarrow$ minimizing the predictive error is to maximize fit in the sample

Prediction Error

Challenge:

- The goal of machine learning is *out of sample* prediction
- Minimize the prediction error outside of the sample
- OLS designed to minimize inside the sample
- Predicting well in sample doesn't mean that it would work outside
- There are estimators that work very well in sample but very badly outside (Overfit) more on this later

Train and Test Samples

- Problem with OLS (and other estimators)
  - Minimize the prediction error outside of the sample
  - OLS designed to minimize inside the sample
- A workaround: split the data
  - Training sample: to build/estimate/train the model
  - Test sample: to evaluate its performance

Predicting House Prices in R

Example: Predicting House Prices in R
(switch to `RStudio`)

Review & Next Steps

- Mean Square Error
- Prediction Error
- Train and Test Samples
- Example in `R`
- Intro to `Git(Hub)`

- **Next Class:** Big Data intro, OLS Numerical Properties Computation.

- Questions? Questions about software?

Further Readings

- Davidson, R., & MacKinnon, J. G. (2004). Econometric theory and methods (Vol. 5). New York: Oxford University Press.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). An introduction to statistical learning (Vol. 112, p. 18). New York: springer.
- Friedman, J., Hastie, T., & Tibshirani, R. (2001). The elements of statistical learning (Vol. 1, No. 10). New York: Springer series in statistics.
- `Git` tutorials from BDEEP group at NCSA. Mimeo.
- `Git` tutorial from Prof. Grant McDermott.