L04.01 Bias / Variance

Top of Form

Bottom of Form

**Objectives**

Bias, Variance, and Regularization in Linear Regression: Lasso, Ridge and Elastic Net

Difference and uses

**Recommended Reading**

--------------------------L04--R4-------------------------

--General topic of this module is to introduce concepts and techniques applicable to any M/L algorithm.

--We focus on Bias and Variance and regularization/cross validation techniques to reduce variance.

--Chapter 2 Section 2.1 and 2.2 from ISLR

--https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf

--https://towardsdatascience.com/bias-variance-and-regularization-in-linear-regression-lasso-ridge-and-elastic-net-8bf81991d0c5

--https://medium.com/@jayeshbahire/lasso-ridge-and-elastic-net-regularization-4807897cb722

--https://stackoverflow.com/questions/27801130/extracting-coefficient-variable-names-from-glmnet-into-a-data-frame

--https://www.analyticsvidhya.com/blog/2015/02/avoid-over-fitting-regularization/

--https://www.theanalysisfactor.com/r-tutorial-glm1/

--https://www.r-bloggers.com/how-and-when-ridge-regression-with-glmnet/

-------------------------End of Reading---------------------------

**Lecture Notes:L04 -- Bias Variance Tradeoff**

[L04-BiasVarianceCV-02.pdf](https://bbhosted.cuny.edu/bbcswebdav/pid-49455053-dt-content-rid-393271767_1/xid-393271767_1)

Instructor notes

* Readings and Resources

Top of Form

Bottom of Form

**Bias - Variance**

[bias-variance.pdf](https://bbhosted.cuny.edu/bbcswebdav/pid-49455056-dt-content-rid-393271768_1/xid-393271768_1)

[**Implementation --Regularization -- L1,L2 and Hybriid (elasticnet)**](https://bbhosted.cuny.edu/bbcswebdav/pid-49455057-dt-content-rid-393271809_1/xid-393271809_1)

[**A Few Useful Things to know about machine learning**](https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf)

[cacm12.pdf](https://bbhosted.cuny.edu/bbcswebdav/pid-49455058-dt-content-rid-393271769_1/xid-393271769_1)

[**Lasso, Ridge and Elastic Net Regularization**](https://medium.com/@jayeshbahire/lasso-ridge-and-elastic-net-regularization-4807897cb722)

"Regularization techniques in Generalized Linear Models (GLM) are used during a modeling process for many reasons"

[**Bias, Variance, and Regularization in Linear Regression: Lasso, Ridge, and Elastic Net — Differences and uses**](https://towardsdatascience.com/bias-variance-and-regularization-in-linear-regression-lasso-ridge-and-elastic-net-8bf81991d0c5)

"Regression is an incredibly popular and common machine learning technique. Often the starting point in learning machine learning, linear regression is an intuitive algorithm for easy-to-understand problems. It can generally be used whenever you’re trying to predict a continuous variable (a variable that can take any value in some numeric range), linear regressions and its relatives are often strong options, and are almost always the best place to start."

**Review Questions and Practice Problems**

1. Can BIAS and VARIANCE both be reduced simultaneously?

2. Resampling and Repeating the exercise , which of this is most likely to change? Bias or Variance?

3.  The Business Unit dictates your DSC team to reduce BIAS at all costs. Which of the regularization technique will you use? LASSO, RIDGE or ElasticNet?

4. In one of your DSC exercise you experienced a very high variance. What would you do to manage that variance to an acceptable level?

5. When are you most likely to experience BIAS when the model is under-fitting or over-fitting?

6. If the business unit demands that you seek to cut the number of features, what penalization method will you prescribe?