

Statistical Learning and Data Science

Course Description:

This short course will focus on statistical learning for data science. A mix of key methodology and research-oriented topics will be covered for statistical learning and data science. The scope will be both methodological with useful applications. The key methodology of statistical learning will be introduced, including regression, classification, graphical model, dimension reduction, and unsupervised learning. The course will also cover a few research topics in discriminant analysis, multi-response regression, interface between machine learning and experimental design, graphical model, and spatial-temporal modeling.

Textbook: T. Hastie, R. Tibshirani and J. Friedman (2009). *The Elements of Statistical Learning*. 2nd edition, Springer-Verlag.

The book is available at <http://www-stat.stanford.edu/~tibs/ElemStatLearn/>

Other required material: Statistics software such as R and MATLAB

Topics:

1. Introduction to data science.
2. Overview of statistical learning: supervised versus unsupervised learning
3. Methods for supervised learning
4. Methods for unsupervised learning
5. Advanced topics in model-based clustering
6. Advanced topics in interface between machine learning and experimental design

References:

1. T. Hastie, R. Tibshirani and J. Friedman (2009). *The Elements of Statistical Learning*, 2nd Edition. Springer-Verlag.
2. Richard A. Berk (2010). *Statistical Learning from a Regression Perspective*. Springer-Verlag.
3. Santner, T. J., Williams, B. J., and Notz, W. I. (2003). *The Design and Analysis of Computer Experiments*. Springer-Verlag.