STA 141A
Fundamentals of Statistical Data Science

Fall 2016

Instructor: Debashis Paul

Lecture 17

Elements of statistical learning

- The last four lectures will focus on various methods of statistical learning, including regression, classification and clustering, as well as data-driven inferential procedures based on the principles of resampling and cross-validation.
- The topics covered in these last four lectures (we shall use materials from Chapters 1-6 and 10 of the James et al., 2013 as reference) will form a major component of the syllabus for the final exam.
- Reference: James, G., Witten, D., Hastie, T. and Tibshirani (2013). *"An Introduction to Statistical Learning with Applications in R"*. Springer.

Linear regression

- Linear regression constitutes one of the simplest examples of supervised learning, where we have a vector of *covariates* or *predictors*, denoted by X, and a *response variable*, denoted by Y.
- The data (X_i, Y_i), i=1,…,n, is assumed to follow a *linear regression model*, if the following description holds:
- **$E(Y|X = x)$ is a linear function of x** for arbitrary x.
- This relationship can also be expressed as saying the *residual term* in the corresponding linear regression model has mean zero (conditional on X).
- Typically, we use the *least squares method* for estimating the regression coefficient in the linear regression model.
- The estimated *regression coefficient* can be expressed as a **linear function of $Y_1, \ldots, Y_n$**

Measure of quality of estimates

- At the population level, the quality of an estimator of a parameter may be measured by the expected squared difference between the estimator and the parameter. If the parameter is vector valued, we replace the squared difference by squared l_2 norm of the difference. This object is referred to as the *Mean Squared Error (MSE)* of the estimator.
- A basic decomposition is:

  MSE(estimator) = Variance(estimator) + (Bias(estimator))^2

- We shall refer to the above equation as **Bias-Variance decomposition of MSE**.
- For typical nonparametric procedures, such as kernel-based nonparametric regression, k-means regression, etc., an optimal choice of tuning parameters depends on balancing between the variance and bias of the estimate, a phenomenon referred to as *Bias-Variance trade-off*.