Overview of Statistical Learning

Seung Hyun Sung

11/4/2021

Contents

> In essence, statistical learning refers to a set of approaches for estimating f. In this chapter we outline some of the key theortical concepts that arise in estimating f, as well as tools for evaluating the estimates obtained. — pg 17 *An Introduction to Statistical Learning with Apllications in R.*

1A Introduction to Regression Models

Depending on the family of Regression model and its complexity of $f(x)$, we may be able to understand how each compnent X_j affects Y, in what particular fashion. Depending on the task (target variable Y) will vary in weight of importance between the interpretation and accuracy. Hence, the phase we are with predicting or defining the task it is important to take this in account before designing and selecting a model.

The notation for ideal Regression function:

$$f(x) = E(Y|X = x)$$

Immediately it emphasis that this regression function gives conditional expectation of $Y|X$. Is there an ideal $f(X)$?

The ideal regression function means the expected value (average) of Y given X. In section 1B we will explore more on this conditional expectation, its useful properties, and geometric interpretation.

For example, if X had three components $x \in R^3$ It is going to be a conditional expectation of Y given three particular instances of these three components of X.

$$f(x) = f(x_1, x_2, x_3) = E(Y|X_1 = x_1, X_2 = x_2, X_3 = x_3)$$

The question given to the function is that at particular point X with three coordinates, X_1, X_2, X_3, what is good value for the function at that point of instances.

Meaning of the ideal or optimal regression function:

- Conditional Average - $E(Y)$ would be the averages of $Y's$ at these coordinates, and the regression function would do that at all points in the plane.

- Ideal means with regard to a loss function, the particular choice of the function $f(x)$ will minimise the sum of squared errors. I.e.

$$f(x) = E(Y|X = x)\, is\, the\, function\, that\, minimises\, E((Y - g(X))^2|X = x)\, over\, all\, functions\, g\, at\, all\, points\, X = x$$

- At each point X , there will be mistakes.

$$E((Y - \hat{f}(X))^2|X = x) = (f(x) - \hat{f}(x))^2 + Var(\epsilon)$$