**What is statistical learning?**

Statistical learning refers to coming up with an unknown function , that helps predict the response using a set of predictor variables . The very general form is as follows

* is the response variable we are trying to predict
* represents the systematic information that provides about . It’s the machine learning model which would be used for prediction. in general, is unknown.
* is the random error term
* Which is independent of and has mean 0.
* It is not captured by the function
* It captures measurement error and other discrepancies.
* In general, which is the estimate of , will not be perfect, so there will be some inaccuracy or error and this error is referred to as reducible error. It’s reducible, because it can be reduced by using the most appropriate statistical learning technique for estimating .
* Even if we come up with the most perfect estimate for , our prediction would still have some error, because is also a function of which cannot be predicted using . This is known as the irreducible error.

**Why do we need to estimate ?**

There are 2 main reasons:

1. Prediction
2. Inference

|  |  |
| --- | --- |
| Prediction | Inference |
| Prediction is, coming up with a technique for estimating with the aim of minimizing reducible error.   * represents the estimate for * represents the resultant prediction for | In inference, we want to understand the relationship between predictor X and response Y, or more specifically, understand how Y changes as a function of (X1, X2 ,…., Xn) |
| If prediction is the goal, non-linear models should be the choice, since they have higher prediction accuracy. | If inference is the goal, linear models might be a good choice, as linear models are more interpretable. |
| Example   * + - Predict response for a marketing campaign [predict if an individual would respond positively to a mail] | Example   * + - Which predictors are associated with the response?     - How much increase in sales (response) is associated with a given media channel (predictor)?     - What effect will change of price (predictor) have on sales (response) of a model? |

**Prediction vs. Inference**