* Which predictors are associated with the response. Only a small set of predictors are typically associated with a response Y.
* What is the relationship between the response Y and each predictor.
* Can the relationship between Y and each predictor be described as linear etc or is the relationship more complicated

Parametric vs Non-Parametric

Parametric involves a two-step approach - We make an assumption about the functional form or shape of f e.g a simple assumption is that f is linear. The second step in parametric approach is to fit the training data by estimating the parameters of the linear model. This approach is called parametric because it reduces the problem of estimating the model to the estimation of parameters. This is much simpler.

Non-Parametric Methods - These do not make any explicit assumptions about the function form of f. Instead they seek an estimate of f that gets as close to the datapoints as possible without being too rough or wiggly. They require a large number of observations in order to obtain an accurate estimate of f.

Assessing Model Accuracy

We need to quantify the extent to which the predicted response value for a given observation is close to the true response value for that observation. In regression, the most commonly used measure is the mean square error (MSE) given by:

When the value of Y is qualitative and not quantitative, the most common approach for quantifying the accuracy of our estimate f is the training error rate, the proportion of mistakes that are made if we apply our estimate f to the training observations i.e C:\Users\OONYIM~1\AppData\Local\Temp\msohtmlclip1\01\clip_image002.png

A good classifier is one which the error rate is small. In other words we should simply assign a test observation with predictor vector to the class j for which p(Y=y|x=) is largest.

Linear Regression

Linear regression assumes that there is a linear relationship between X and Y. I.E , and are coefficients or parameters. Training data is used to estimate and fit both parameters and then can be used to estimate future sales on a particular value of TV advertising by computing,

, where y’ is a prediction of y on the basis of X=x.