- Linear regression assumptions (i)Additive assumption $\rightarrow$ Effect of changes in a predictor X_j on the response Y is independent of other predictors. (Solution take intercept of two predictors) (ii)Linear assumption $\rightarrow$ states that the change in the response Y due to change in X_j is constant, regardless of the value of X_j (iii) Error terms, ϵ_1, ϵ_2,...,ϵ_n are uncorrelated.(If we plot error terms there shouldn't be any pattern) (iv) Error terms have a constant variance (heteroscedasticity $\rightarrow$ non-constant variances in the errors, makes funnel shape in residual plot. one possible solution is to transform the response Y using a concave function such as $\log Y$ or $\sqrt{Y}$)
- RSS(ResidualSumOfSquares) $= (y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_1)^2 + (y_2 - \hat{\beta}_0 - \hat{\beta}_1 x_2)^2 + \cdots + (y_n - \hat{\beta}_0 - \hat{\beta}_1 x_n)^2$
- RSE (measure of the lack of fit of the model)= $\sqrt{\frac{1}{n-p-1} RSS} = \sqrt{\frac{1}{n-p-1} \sum\limits_{i=1}^{n} (y_i - \hat{y}_i)^2}$
- TSS(TotalSumOfSquares)= $\sum_{i=1}^{n} (y_i - \bar{y})^2$, And ... $R^2 = \frac{TSS-RSS}{TSS}$
- MeanSquaredError (values closer to zero are better)= $\frac{1}{n} \sum\limits_{i=1}^{n} (y_i - \hat{f}(x))^2$, $\rightarrow \hat{f}(x_i)$ is the prediction that $\hat{f}$ gives for the i^{th} observation
- Standard error $SE(\hat{\beta}_0)^2 = \sigma^2 \left[\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2} \right]$ and $SE(\hat{\beta}_1)^2 = \frac{\sigma^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$
- F-Statistics = $\frac{(TSS-RSS)/p}{RSS/(n-p-1)}$ ...(p predictors,n observations.When n is large,F-statistic just above 1 might provide evidence against H_0. If n is small a larger F-statistic is needed to reject H_0)
- If Null Hypothesis is correct $E\{TSS - RSS/p)\} = \sigma^2$, If linear model assumptions are correct $E\{RSS/(n-p-1)\} = \sigma^2$, So when there is no relationship between the response and predictors, F-statistic ≈ 1
- Cor(X, Y) = $\frac{\sum\limits_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum\limits_{i=1}^{n} (x_i - \bar{x})^2} \sqrt{\sum\limits_{i=1}^{n} (y_i - \bar{y})^2}}$ ...In multiple linear regression $Cor(Y, \hat{Y})^2 = R^2$
- multicollinearity $\rightarrow$ Collinearity can exist between three or more variables even if no pair of variables has a particularly high correlation. Variance inflation factor (VIF) assess multi-collinearity. $VIF(\hat{\beta}_j) = \frac{1}{1 - R^2_{X_j|X_{-j}}}$. VIF > 5 - 10 Suggest strong collinearity.
- Degree of freedom = n - k -1
- KNN $\rightarrow Pr(Y = j|X = x_0) = \frac{1}{K} \sum\limits_{i \in N_0} I(y_i = j)$ (Given a +ve int K & observation x_0, KNN classifier first identifies the K points that are closest to x_0 (represented by N_0), then estimates the conditional probability for class j as the fraction of points in N_0 whose response values equal j
- KNN 2^{nd} form $\hat{f}(x_0) = \frac{1}{k} \sum\limits_{x_i \epsilon N_0} y_i$
- SSR(SumOfSquaredDueToRegression) = $\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$, SSE(SumOfSquaredDueToError) = $\sum_{i=1}^{n} (y_i - \hat{y}_i)^2$, $SST = SSR + SSE = \sum_{i=1}^{n} (y_i - \bar{y})^2$ and $R^2 = \frac{SSR}{SST}$ =(Proportion of the variation of y being explained by the variation in x)
- HUNT $\rightarrow$ (I) The initial tree contains a single node with class label that has majority of the out come (II) If ail the records in D_t belong to the same class y_t, then t is a leaf node labeled as y_t (III) Else partition the records into smaller subsets that yields highest Info gain $\Delta_{info} = I(parent) - \sum_{j=1}^{k} \frac{N(v_j)}{n} I(v_j)$. N is total records at parent Node, k is number of attributes, $N(v_j)$ is number of records associated with child node v_j. I(.) = Impurity of a node = $-\sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t)$ { ...c is number of classes , $p(i|t)$ fraction of records having class i at a node t} .This algorithm is then recursively applied to each child node.
- GiniIndex $\rightarrow \sum_{i=0}^{c-1} p(i|t)^2$ {Can be used instead of entropy above } .Exam-Tan-P160-Sum-prob
- bagging $\rightarrow \hat{f}(x) = \frac{1}{B} \sum_{b=1}^{B} \hat{f}^{*b}(x)$ {B = bootstrapped training data sets.$\hat{f}^{*b}(x)$ = Predicton when model trained on b^{th} dataset, and finally average all the predictions }
- ConfusionMatrix $\rightarrow Sensitivity = \frac{TruePositive}{TruePositive+FalseNegative}, Specificity = \frac{TrueNegative}{TrueNegative+FalsePositive}$