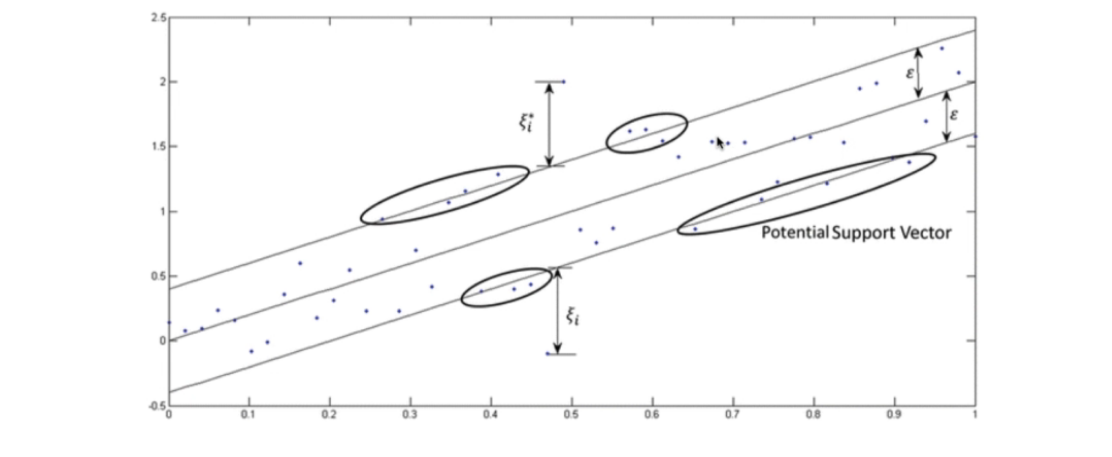
Non-Linear regression:

Types of non-linear regression models

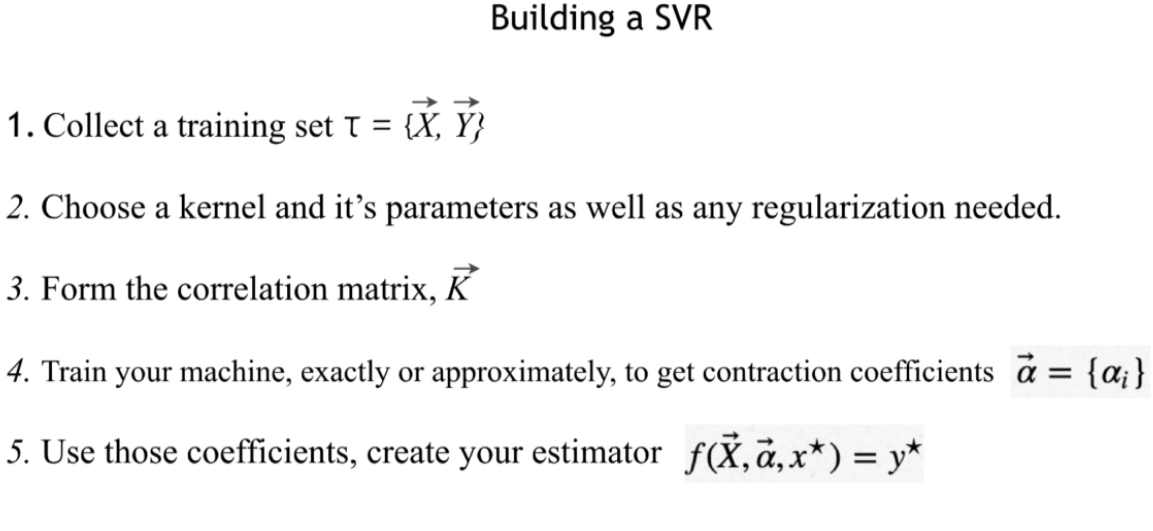
* SVR (Support vector regression)
* Decision tree regression
* Random forest regression.

**SVR (Support Vector Regression)**

* Supports linear and non-linear regression.
* Performs linear regression in higher dimension.



**Steps to perform SVR:**



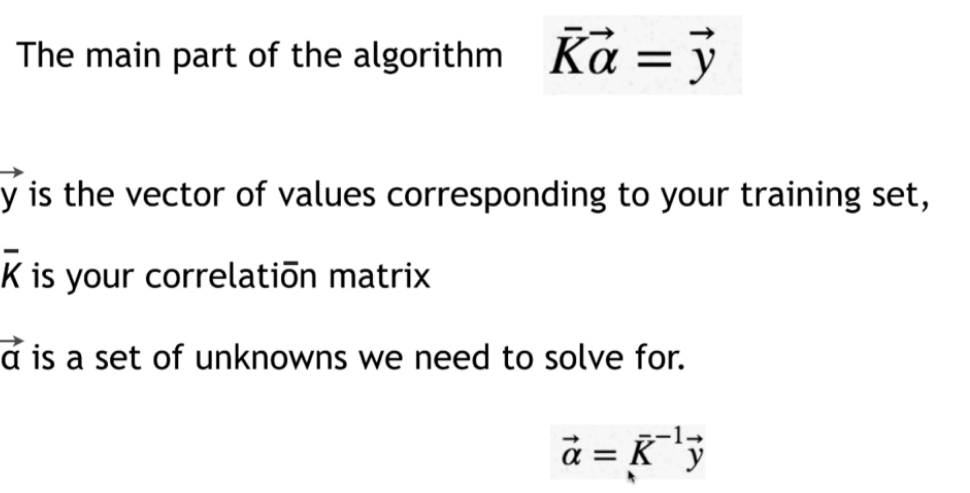
Correlation matrix formation will be taken care by the SVR package.

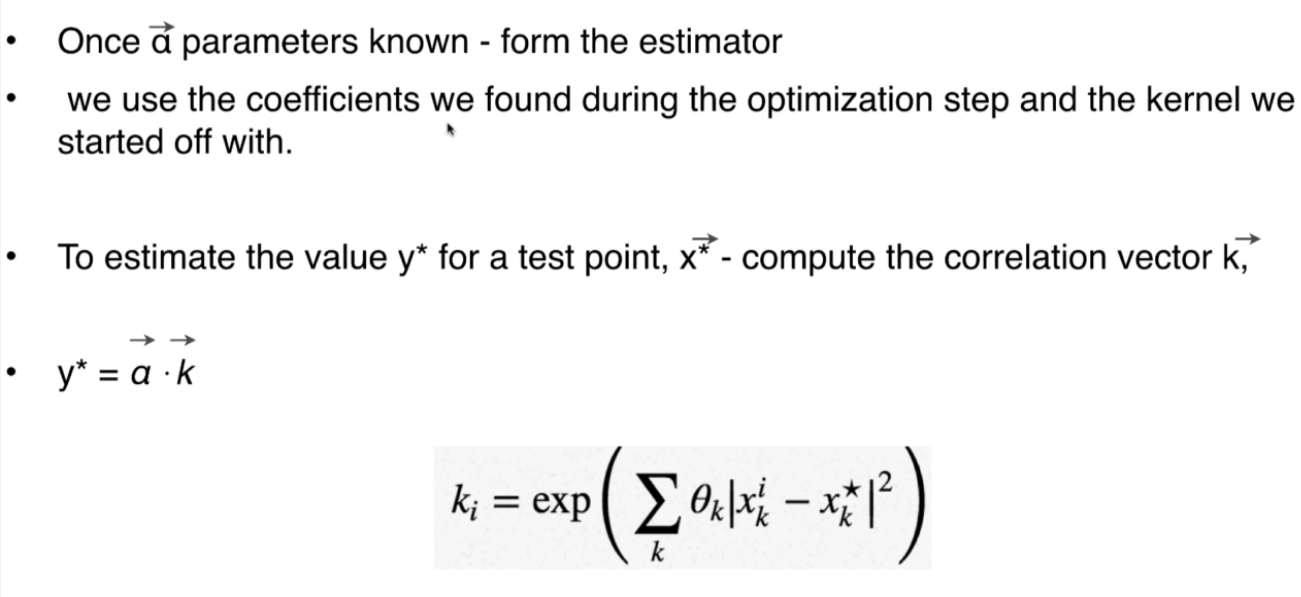
**Choosing kernels:**

* Gaussian – Default kernel for SVR.

**Regularization:**

* Noise - Reducing noise



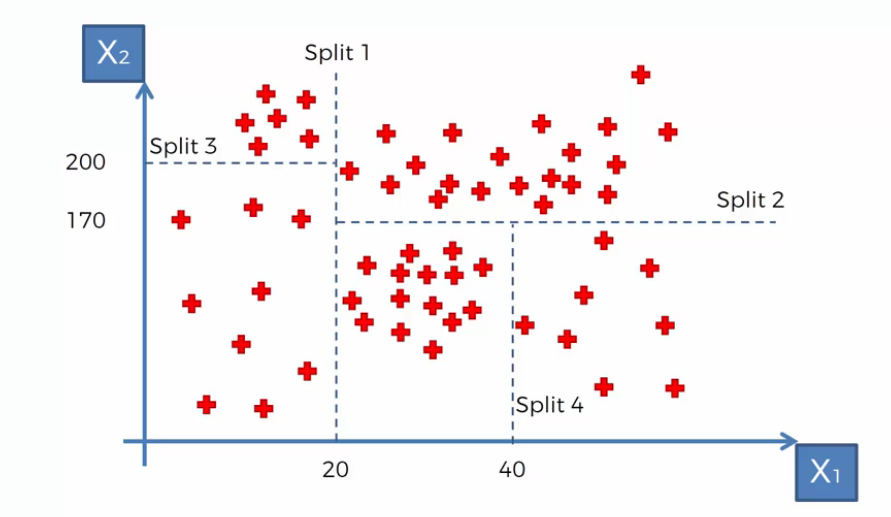


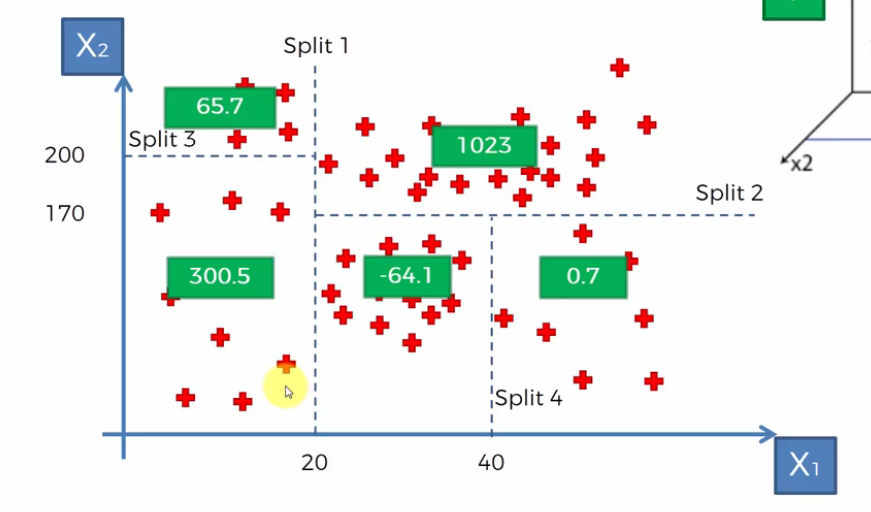
**Goal of SVR:**

In linear regression, we try to reduce the error between predicted value and actual value but in SVR we must make sure the error value doesn’t cross the threshold value.

**Decision Tree Regression:**

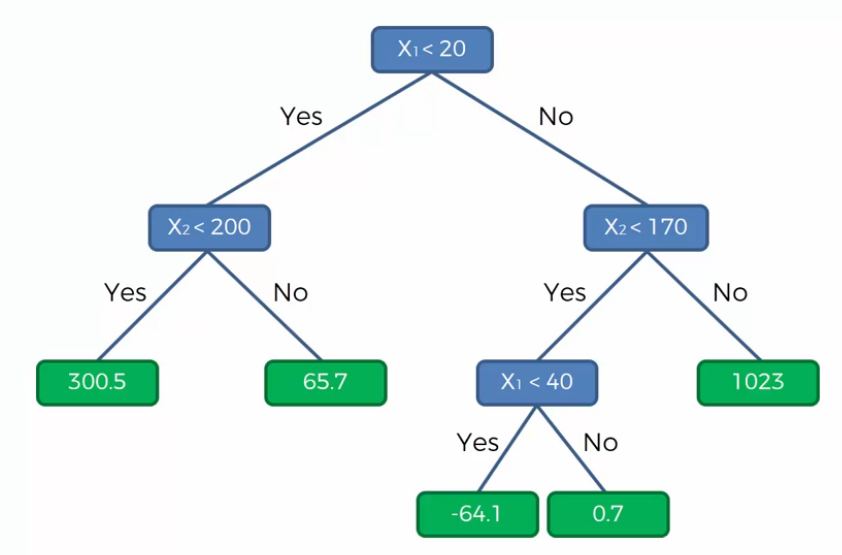
* Plot the scatter plot of independent variable (any number of independent variable).
* Algorithm tries to split your scatter plot based on algorithm logic based on information entropy.
* Splitting will increase the value of information.
* Algorithm stops when it cannot add more information or has less than 5% of total points.





Green box denote the mean value of the split segment.

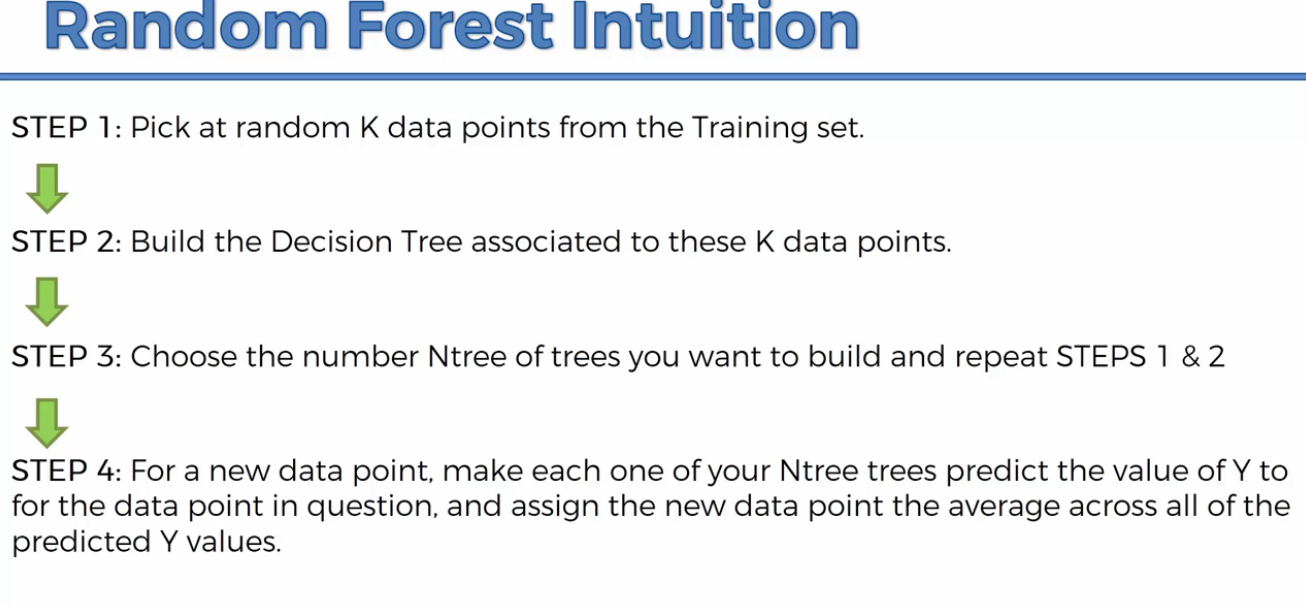
**Final tree:**



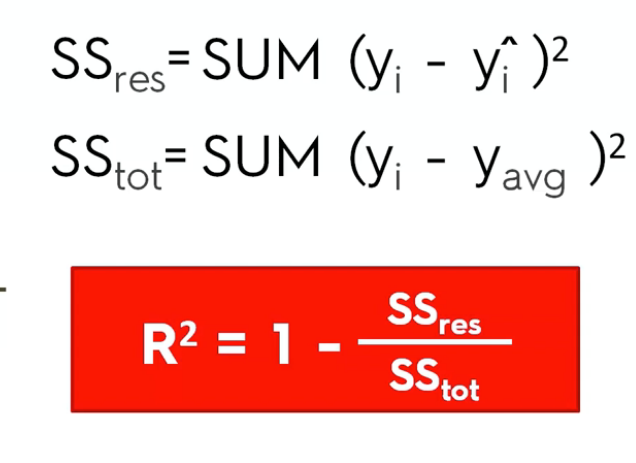
**Random forest Algorithm in Regression tree:**

* Version of ensemble learning (Same algorithms put together multiple times to improve the prediction value).
* Ensemble algorithms are more stable because any change in data set will affect only one tree not the whole forest.

**Algorithm:**



**Evaluation of Regression models:**

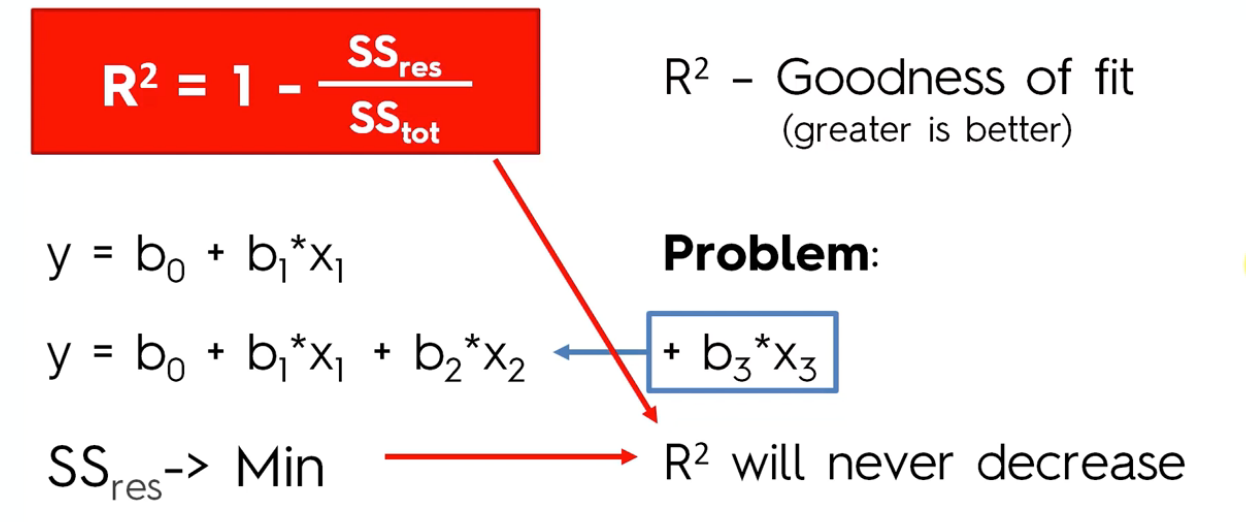
**R- Squared: To find the goodness of fit.**

Res = sum of squares of residual; tot = sum of squares of average.

Our aim in designing the model is to minimize the sum of squares of residual.

SSres = 0, in that case R^2 = 1 which means the linear line goes through the plot. R^2 closer to 1 has better prediction value. R^2 value can also be negative.

**Adjusted R square:**

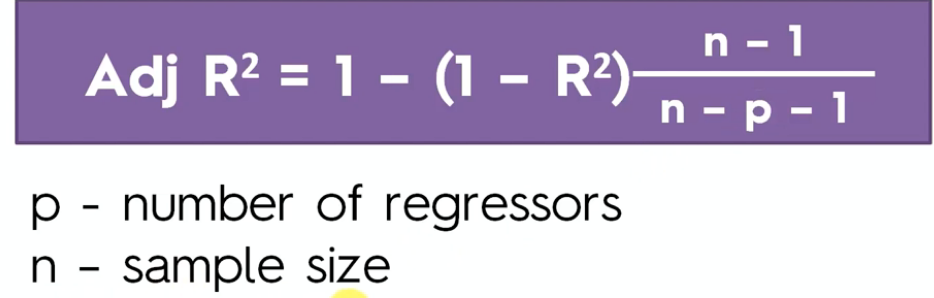


**R^2 will either increase due to introduction of new variable or be equal to the old R^2 value.**

**It will never decrease.**

**Problem with R^2 is Adding the new variable will not show whether the new variable is helping your model or not.**

**To overcome this goodness of fit, we have adjusted R^2.**



**Adjusted R^2 has a penalization factor. It penalizes for adding independent variable that don’t help your model.**

* **Adding more regressor (Independent variables), the adjusted R^square will decrease.**
* **When normal R^2 increases, then adjusted R^square increases.**

**If your variable is not helping your model then there is a significance increase in R^2 and the penalization factor of p will drive adj r^2 down.**

**On the other hand, your variable helping your model, then increase in R^2 will be substantial, and it will overwhelm the penalization factor.**