Trees Versus Linear Models

As discuss in Decision Trees both of the **Regression** and **classification** trees are widely different fromMultiple Linear Regression ,Logistic Regression,Linear Discriminant Analysis and other **linear models**.

For **Linear Regression** it assumes a model of form :

$$f(X) = \beta_0 + \sum_{j=1}^{p} X_j \beta_j$$

Where the **Regression Trees** assumes a model of form :

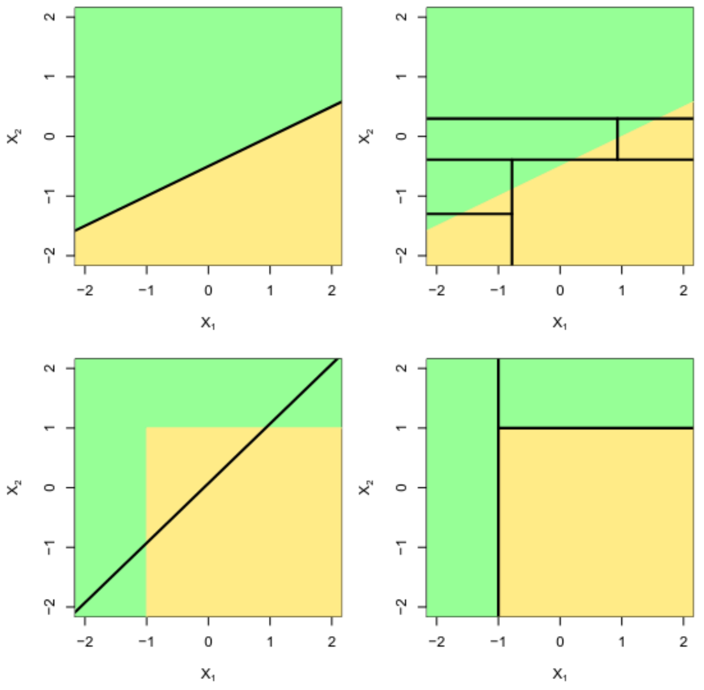
$$f(X) = \sum_{m=1}^{M} c_m . 1(x \in R_m)$$

- Where $R_1 \dots, R_m$ represent partition in the feature space.

Now the question : **Which model is better?**
It depends on the problem at hand, if the **relationship** between the **predictors** and Response is linear-ish, the classical **linear models** will outperform **tree based models**.

If the is a highly non linear and complex relationship between the **features** Response then decision trees may outperform classical models, it can be noticed that the **regression tree** form is quite similar to Regression Splines but simpler since there is no **knot** tuning and fitting.

Also tress maybe be preferred for the sake of **interpretability*** and **visualization**.



But due to the **low bias- high variance** of fully grown tress **ensemble learning** target that by sequentially correcting multiple tress which reduce the high variance, and the non-linearity allow them to be a powerful tool for prediction accuracy, finally the **feature selecting** since Decision Trees select the features splits for the most information gain