

# Trees Versus Linear Models

As discuss in [Decision Trees](#) both of the **Regression** and **classification** trees are widely different from [Multiple 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 :

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

- 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 trees maybe be preferred for the sake of **interpretability\*** and **visualization**.

## Advantages and Disadvantages of Trees

### Advantages

- Very easy to explain to non-technical people more than **linear regression**
- Ease of interpretability for it's human like decision making
- No Statistical assumptions
- Little Data preprocessing needed [Feature Scaling](#)
- Can handle qualitative predictors(features) without dummy variables
- Easy to visualize

### Disadvantages

- Deep trees lost their explanability advantage
- Poor with **linear relationships**
- Require extensive tuning
- Overfit easily