Regression Trees

Timothy Currie

Wissenschaftliches Arbeiten
14/06/2024

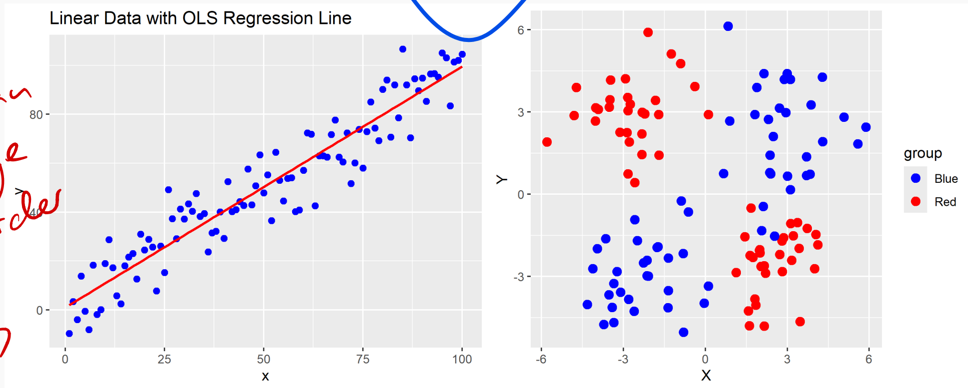
Motivation

- OLS often performs poorly on many kinds of data ~~data~~
- **Example:** Non-linear relationships and interaction effects.

show plot after this bullet point

- In these situations, Regression Trees can be a better tool.
- Basic Regression Trees have lots of limitations so we also discuss pruning and Ensemble methods than can improve results.

Maybe you can change the order



(a) OLS performance

(b) Tree performance

Figure 1: Performance comparison between OLS and Trees

Outline of the Talk

1. Motivation
2. Basics of Trees
3. Difference to OLS
4. Overfitting and Pruning
5. Ensemble methods
6. BART
7. Conclusion
8. References

Basics of Regression Trees and difference to OLS

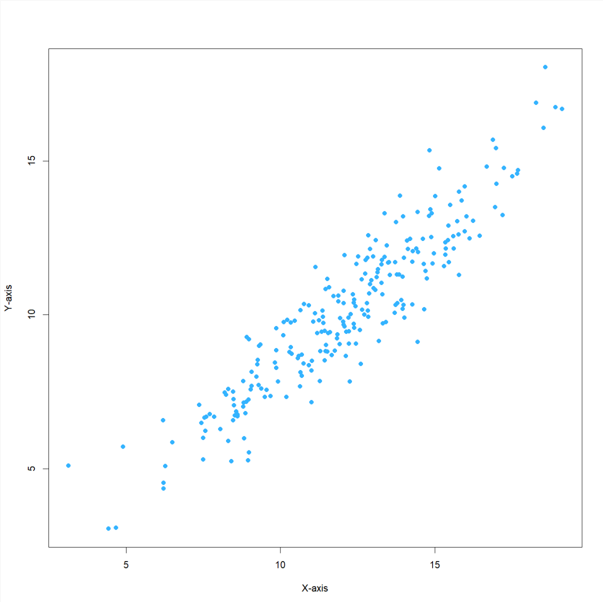
- Regression trees are an alternative to linear Regression
- They split the predictor space into boxes that minimize the RSS given by

$$\sum_{j=1}^{J} \sum_{ic \in R_j} (y_i - \hat{y}_{R_j})^2$$

- Sadly we can't just find these optimal boxes, in practice we use a greedy algorithm **Recursive binary splitting** to find the optimal split at each stage to minimize prediction error.
- Unlike OLS, Trees do not assume a linear relationship between predictors and the response.
- Trees can capture interaction effects naturally.

- Lets find out how these different methods do in a Simulation

Simulation:Linear and Non-Linear Relation between Variables



- How we will Run the simulation:
- Take the data decide parameters
- Regression Trees will have 4 terminal Nodes
- Run models on data, compare Mean Squared Error

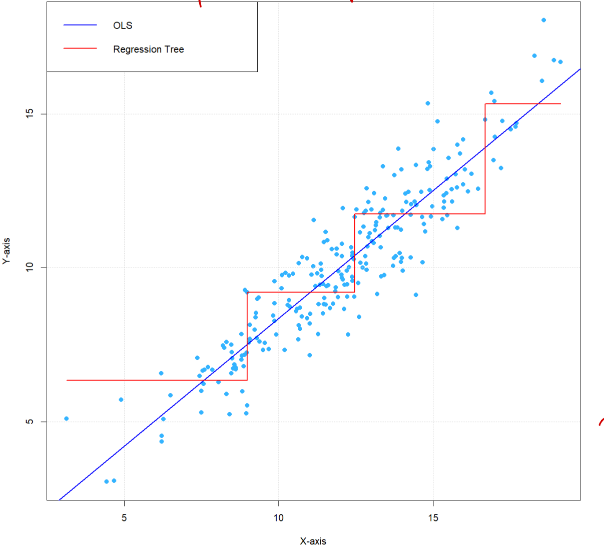
$\Rightarrow$ What is the data-generating process here?

$$y_t = f(x_t) + \varepsilon_t$$

$\hookrightarrow$ $f(x_t)$ linear? non-linear?

Simulation:Linear and Non-Linear Relation between Variables

→ Very nice plot



- How we will Run the simulation:
- Take the data decide parameters
- Regression Trees will have 4 terminal Nodes
- Run models on data, compare Mean Squared Error

- **Results:**
- **MSE** for OLS model: **0.9797**
- **MSE** for regression tree model: **1.3752**

For M Simulations, you should compute and report the average $\overline{MSE}$:

$$\overline{MSE}^{OLS} = \frac{1}{M}\sum_{i=1}^{M} MSE_i^{OLS} \qquad \overline{MSE}^{TREE} = \frac{1}{M}\sum_{i=1}^{M} MSE_i^{TREE}$$

Simulation: Non-linear Data

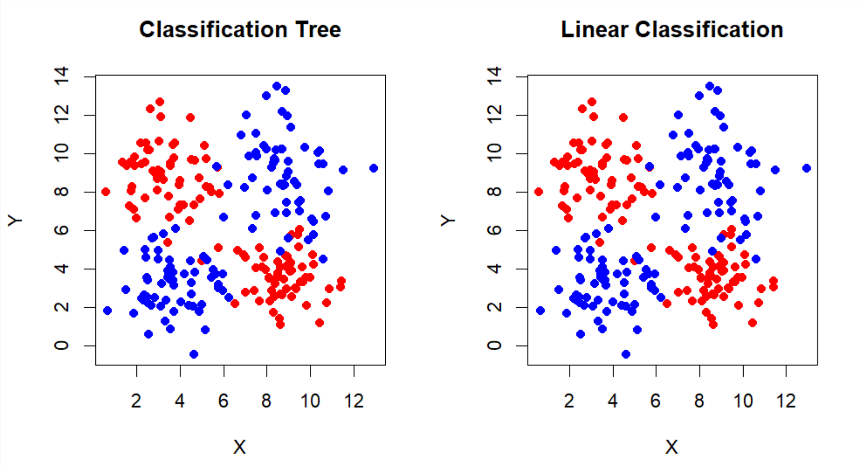


Figure 2: Data with non linear relationship

- How will the two Models Perform here on Non-Linear Relation between Variables?

⇒ DGP?

Simulation: Non linear Data, Model performance

What is the interaction effect?

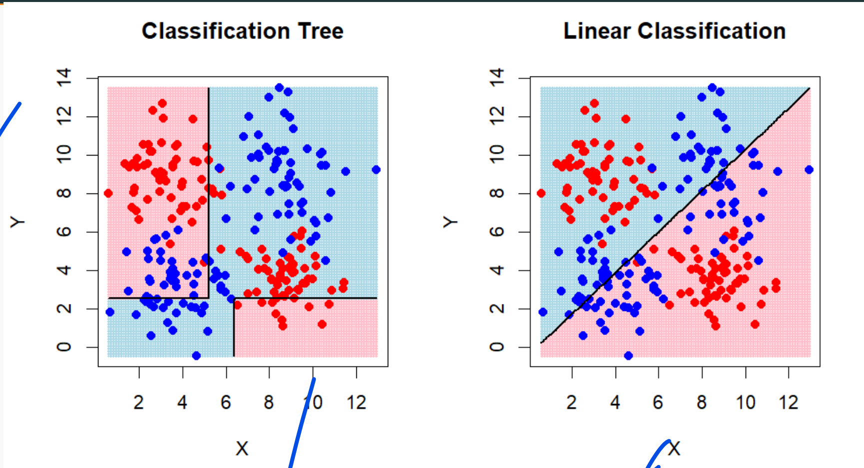


Figure 3: Data with Prediction

you can shift this upwards using the function \vspace (−0.5cm) in Overleaf

- Classification Tree MSE: 0.3542
- Linear Regression MSE: 0.5167
- Here, Classification Error Rate = MSE
- **Reason Trees are better:** Trees naturally capture interaction effect

Two Problems With Regression Trees

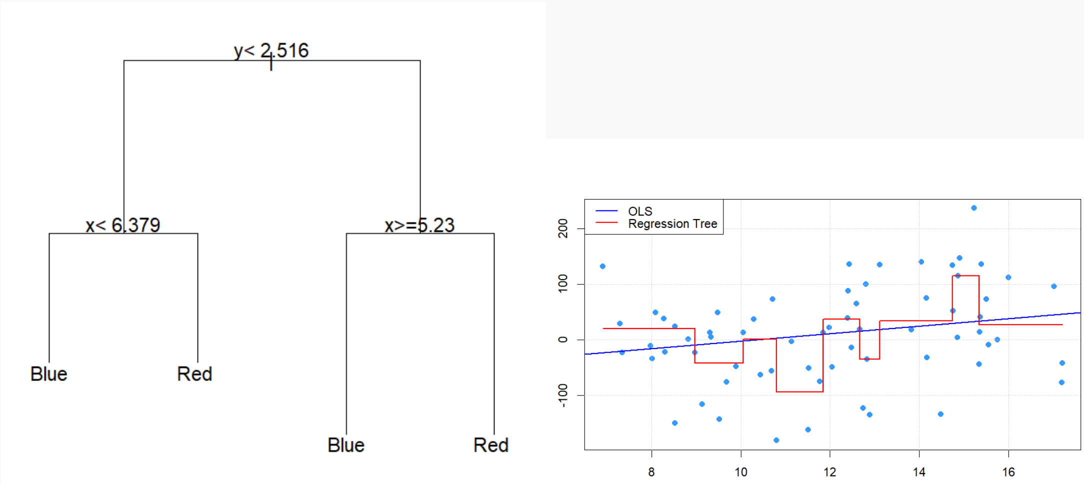


Figure 4: Two Problems that can arise with Trees

- Overfitting and non Optimal Splitting $\Rightarrow$ very nice plot

Pruning

- **Cost complexity pruning** counteracts overfitting by removing non-essential splits.
- Lets us grow a large Tree and then

$$\sum_{m=1}^{|T|} \sum_{i:x_i \in R_m} (y_i - \hat{y}_{R_j})^2 + \underline{\alpha|T|}$$

  - Select a parameter α.
  - For each α, find the subtree that minimizes the cost.
  - Use cross-validation to select the best α.
  - Instead of evaluating a Model on the data we trained it on we evaluate it on a seperate set.
- **Objective:** Achieve a good tradeoff between bias and variance.

$\Rightarrow$ To help your audience maybe add one line: if $\alpha >> 0$ Trees are small/shallow $\rightarrow$ not many splits

Finding the optimal Tree size

Nice plot → Add one line with intuition

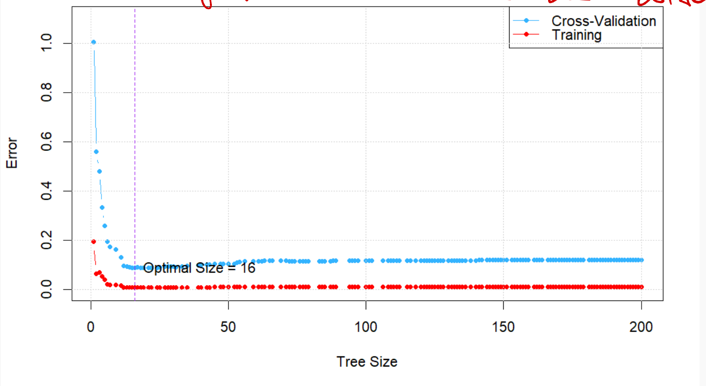


Figure 5: Training and Cross-Validation Error

- Pruning helps against overfitting and improves overall performance
- Original Tree MSE on Test set = 277.5816
- Pruned Tree MSE on Test set = 232.295

⇒ Explain difference between Train & Test

Ensemble methods

- Even with pruning trees often perform worse than linear other ML methods
- Ensemble methods improve results by combining many regression trees. Each one contributes a small part to the overall prediction.
- Each tree can be independent of previous trees (e.g. Random Forests)
- Or can be grown on the residuals of the current fit (e.g. Bayesian Additive Regression Trees (BART))

BART

If you run out of time you can skip this slide

- BART models the response as a sum of many tree-based models plus noise.
- **Model:**

$$Y_i = \sum_{j=1}^{m} g(X_i; T_j, M_j) + \epsilon_i \tag{1}$$

- BART calculates the residuals of the current sum of Trees.
- Then modifies one Tree to decrease the residuals.
- Then take the average over all but the burn-in iterations
- Unlike single trees, BART avoids overfitting by averaging the predictions of many trees.
- BART provides a probabilistic prediction, giving a measure of uncertainty.

Conclusion and Discussion

- Regression trees are powerful for non-linear and interactive effects.
- They are also very easy to interpret.
- Trees require pruning to command overfitting.
- By averaging independent Trees or fitting trees on the residuals ensemble methods can improve results.
- BART is a sophisticated method offering good results in many scenarios.

Very nice presentation

⇒ Add some details: DGP: Slide 4 + 6

⇒ Intuition for results of Plot on Slide 10

Train & Test Samples

References

- Elements of Statistical Learning by Hastie, Tibshirani, and Friedman.
- ...
- ...