

Lecture 13: Regression using decision trees

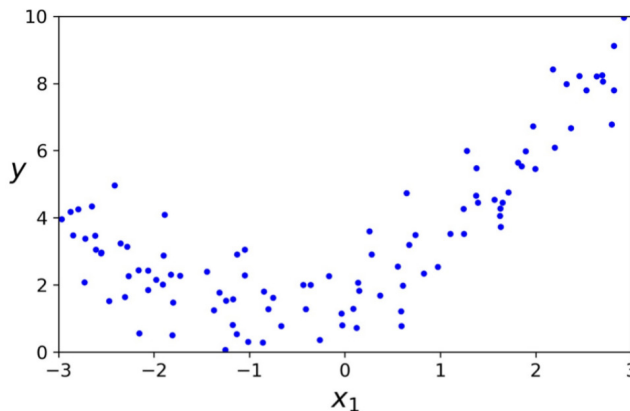
Madhavan Mukund

<https://www.cmi.ac.in/~madhavan>

Data Mining and Machine Learning
August–December 2020

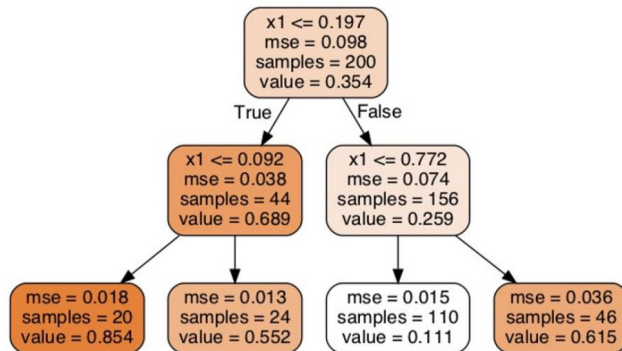
Decision trees for regression

- How do we use decision trees for regression?
- Partition the input into intervals
- For each interval, predict mean value of output, instead of majority class
- Regression tree



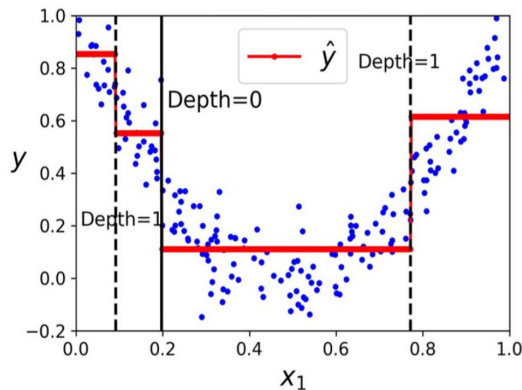
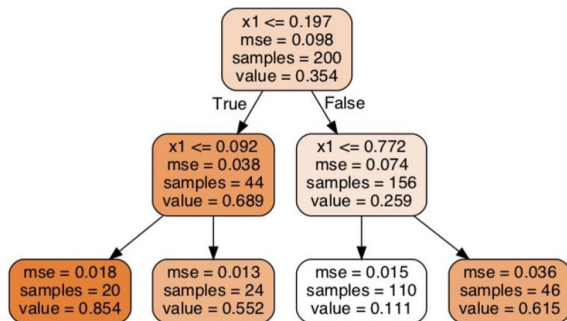
Decision trees for regression

- Regression tree for noisy quadratic centered around $x_1 = 0.5$
- For each node, the output is the mean y value for the current set of points
- Instead of impurity, use mean squared error (MSE) as cost function
- Choose a split that minimizes MSE



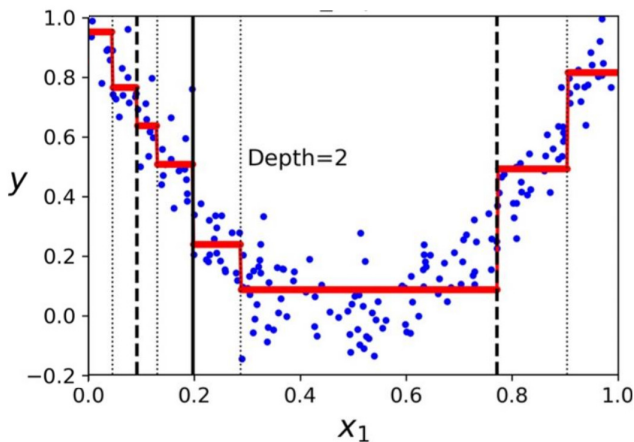
Regression trees

■ Approximation using regression tree



Regression trees

- Extend the regression tree one more level to get a finer approximation
- Set a threshold on MSE to decide when to stop
- **Classification and Regression Trees (CART)**
 - Combined algorithm for both use cases
- Programming libraries typically provide CART implementation



Summary

- Regression trees are decision trees that predict a value rather than a category
- Output value is the mean value of all points in the current node
- Choose query to reduce MSE as much as possible
- Stop when MSE within acceptable tolerance

