Portfolio Question 3.1.P2 - Regression Trees

Daniel Lawson University of Bristol

Portfolio 3.1.P2 (v4.0.0)

- How do we use a tree to perform regression?
  - What is the “natural” loss function?
  - What pitfalls might exist in this approach?

Regression Trees

- Decision trees can also be used for regression
- There is a nice relationship between the loss function of classification, and “mean squared error” in regression

Regression Trees (simple version)

- Regression trees are constructed identically to classification trees
- Decision can follow information or squared loss
- Prediction is the average $\hat{y}_i|\{i \in c_j\} = \bar{y}_{c_j}$ inside the leaf j
- Comparing to classification: if $y \in \{0, 1\}$ this reduces to:

$$R_j = \frac{1}{N_j} \sum_{i \in c_j} (y_i - \hat{y}(x_i))^2 \quad (1)$$

$$= \frac{1}{N_j} \sum_{i \in c_j} \left(y_i^2 - 2 y_i \hat{y}(x_i) + \hat{y}(x_i)^2 \right) \quad (2)$$

- Which since $\hat{y} = \bar{y}$, simplifies to:

$$R = \bar{y} - 2\bar{y}^2 + \bar{y}^2 \quad (3)$$

$$= \bar{y}(1 - \bar{y}) \quad (4)$$

Comparing Regression Trees to Classification Trees

- Similarly the Gini index is:

$$G = \sum_{j=1}^{J} p_j(1 - p_j) \tag{5}$$

$$= \bar{y}(1 - \bar{y}) \tag{6}$$

- So regression trees and classification trees use equivalent loss functions in CART.

General Regression Trees

- Within each decision node (leaf) we fit a model
- This is typically a **constant** model, i.e. the average
- Why?
  - The piecewise constant model fit can be arbitrarily good (number of splits scales with data volume)
  - Making non-constant models consistent is computationally costly
- Why not?
  - Computational cost grows with tree depth
  - Often locally the structure is linear
- Leads to Local Regression Trees [1]
  - Complex if we ensure that the local regressions meet up

[1] Karalic A, "Employing Linear Regression in Regression Tree Leaves" (1992) ECAI-92

Discussion

- Could there be “discontinuities” where the model changes a lot?
- How could we prevent this?