

1.8.6. Tree algorithms: ID3,C4.5,C5.0 and CART

Kazuhei Kouno

ID3

- Ross Quinlan (1986)
- 決定木を出力
- 各独立変数のうち情報利得の最大のものを選択し、木のノードにする

- 枝刈り
 - 未知のデータへ一般化
 - 木の精度が向上

ID3の拡張

- C4.5
 - ID3: 離散値のみ
 - C4.5: 離散値と連続値
- C5.0
 - メモリ使用量が少ない
 - C4.5よりも正確かつ小さな決定木を構築
- CART ← scikit-learnで実装
 - C4.5: 多分岐にも対応・情報エントロピー
 - CART: 2分岐のみ・ジニ係数