Penalized Empirical Risk Minimization and Cross Validation

機械学習の経済学への応用

川田恵介

Penalized Empirical Risk Minimization

- モデル集計と並ぶ人気戦略
  - モデル集計に比べて、理解しやすい予測モデルを得やすい
- 経済学理論においても馴染みのある発想

基本方針: 目的関数の修正

- 目的: Population Risk Minimization
- Empirical Risk Minimization で推定
  - 悪くない発想
- “全てのパラメータ” を決定すると、しばしば深刻な過剰適合が発生
- 対策: 目的関数を” 修正”
  - モデルの複雑さにペナルティーを与える
- 課題: ペナルティーをどう決める

経済学版: 自家用車分配問題

- 目的: Social Welfare Maximization
- (Individual) Utility Maximization を目指した自家用車保有の意思決定
  - 悪くはない
- 過剰な負の外部性（渋滞、汚染、騒音など）が一般に発生
- 目的関数を修正

  - 自動車保有に” 課税”

- “課税” をどう決める？
  - 観察できる情報を用いて頑張る

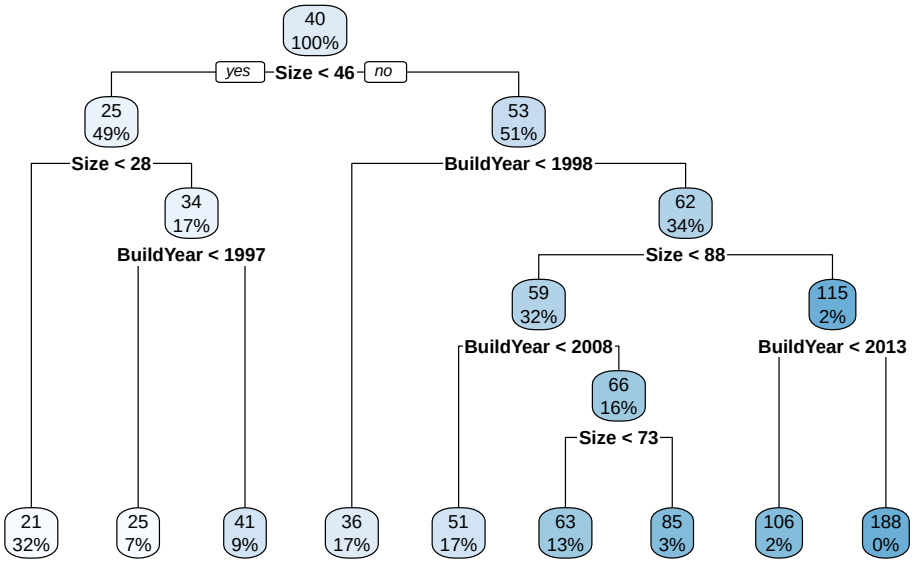
Prune

1. Empirical Risk Minimization の解として、深い予測木を推定
2. Penalized Empirical Risk Minimization の解として、剪定 (サブグループを再結合)

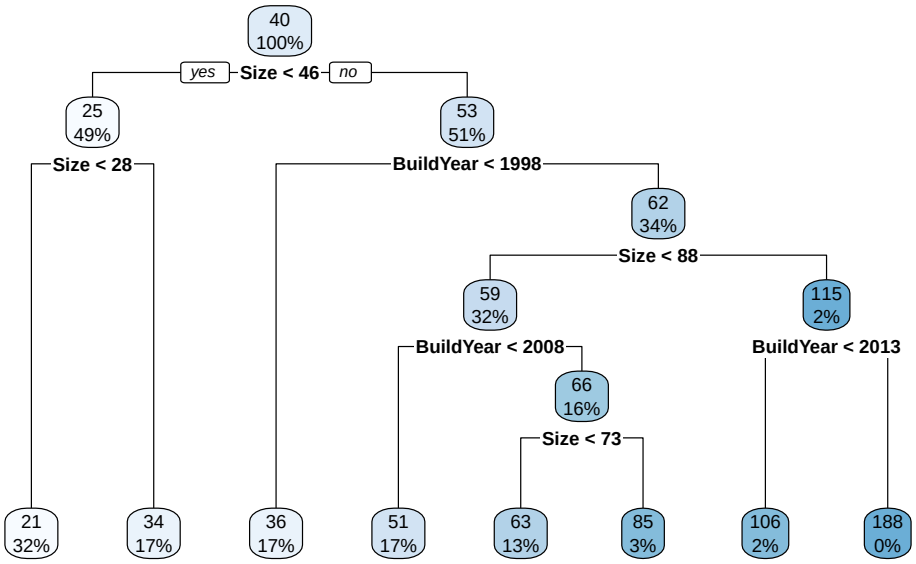
$$EmpiricalRisk + \underbrace{\alpha \times |Number\ of\ SubSample|}_{Penalty}$$

- Empirical Risk 削減に貢献しない分割から、再結合される

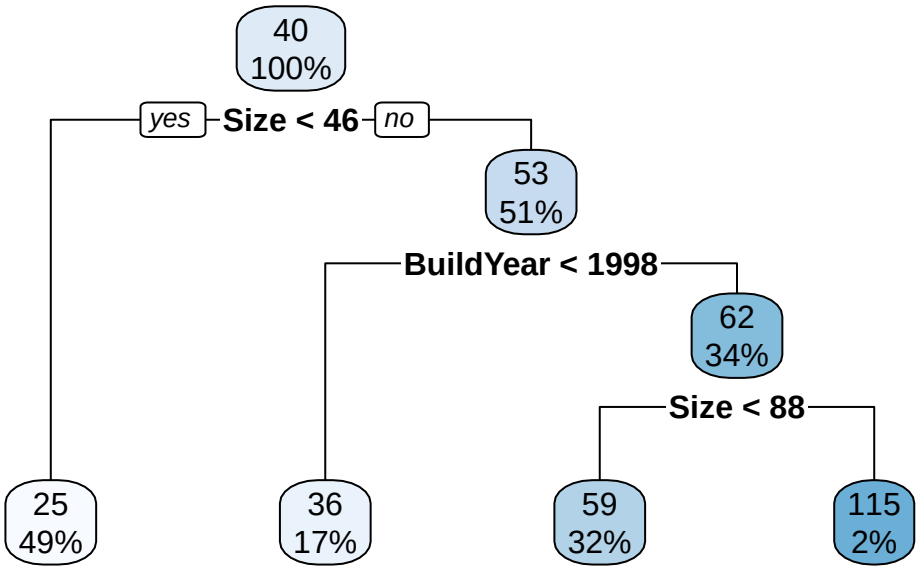
例: alpha = 0.01



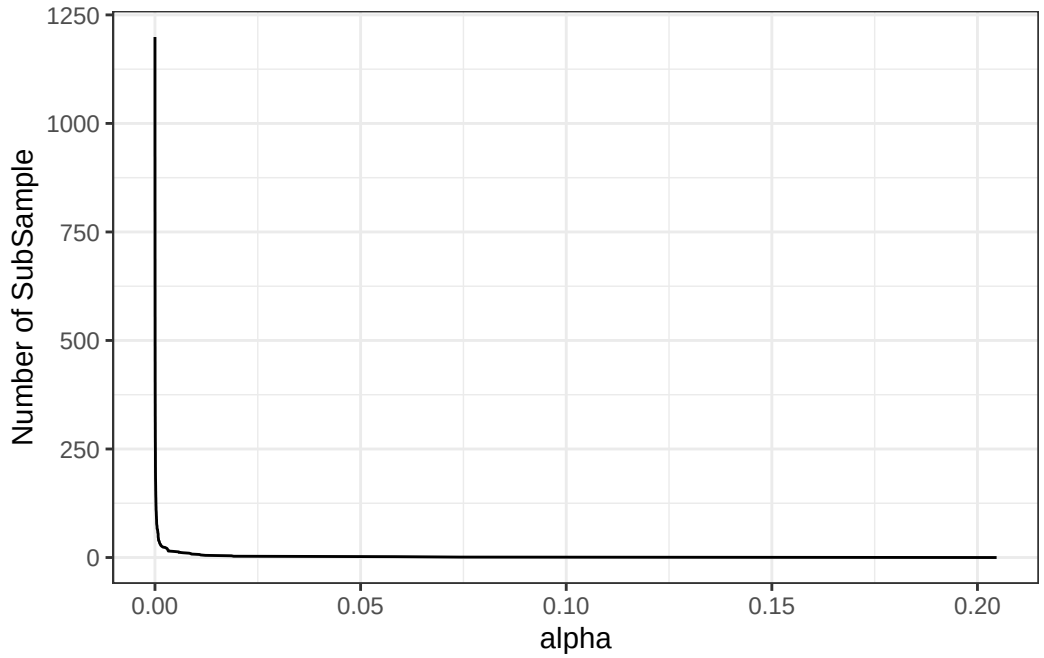
例: alpha = 0.011



例: alpha = 0.02



例: サブグループの数



α の決定問題

- 複数の TuningParameter α を比較し、“最善の値” を決める
  - 最善の値さえ決まれば、あとは全サンプルを用いて、モデルを推定

既習の手法

- 各 α について” ベンチマークモデル” を推定し、評価・比較する
  - ベンチマークモデルと同じデータで評価すると、大問題!!!
- ベンチマークモデル作成データと中間評価データに分割するのは OK だが
  - モデル作成/中間評価のトレオードオフが深刻化

論点

- 以下は予測モデルの評価の優れた推定値

$$\sum (Y_i - f(X_i))^2 = \sum (\mu(X) + \underbrace{u_i - f(X_i)}_{Independent})^2$$

- 事例 i に適用する予測モデルと” 誤差項” が独立であれば OK
  - Test/Training への分割は一つの方法

まとめ

- Penalized Empirical Risk Minimization は、直感的な手法
  - Approximaiton Error を避けために非常に複雑なモデルからスタートし、複雑性への罰則を加えた Empirical Risk Minimization の解として単純化
- 罰則の重さを決めるのが難しい

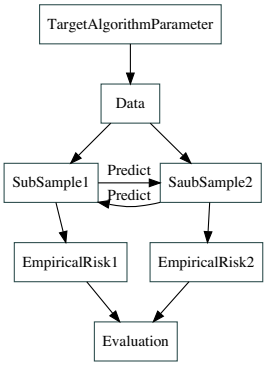
交差推定

- 予測のみならず、比較・因果推論への機械学習の応用においても、非常に重要なテクニック

手順

1. (Training) データをランダムに分割 (2;5;10;20 など)
2. 第 1 サブグループ以外を用いてモデルを推定し、第 1 サブグループの事例について予測値を計算
3. 全てのサブグループについて、繰り返す
4. **全事例**について、誤差項と独立な予測値を得る

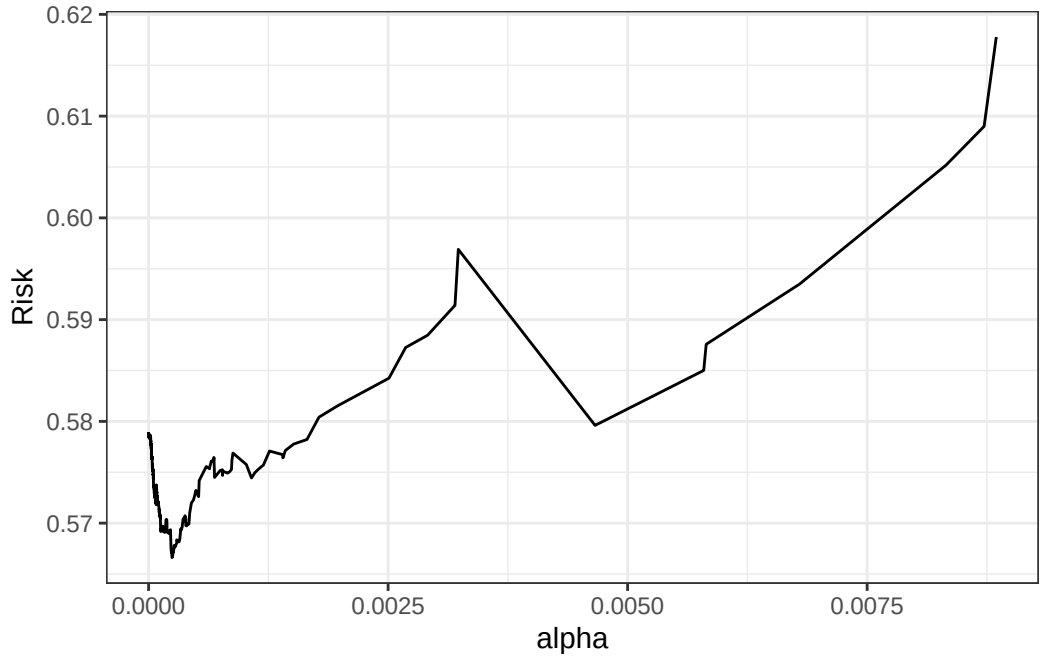
Cross Validation



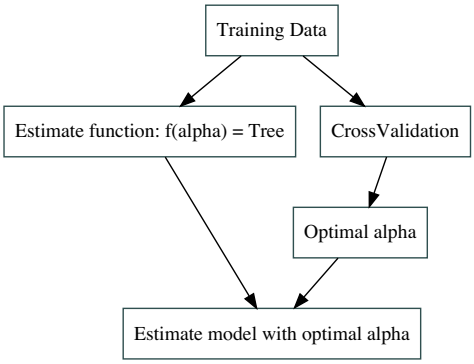
Cross-Validation の利点

- 分割数を増やすと、推定に大量のデータを使える
  - 訓練データ全てを用いたモデルに近づく
- 個々の検証データは少数であり、不安定が大きい
  - 複数の検証結果の平均をとるので、安定
- 分割数を増やすと、計算負荷が劇的に上昇

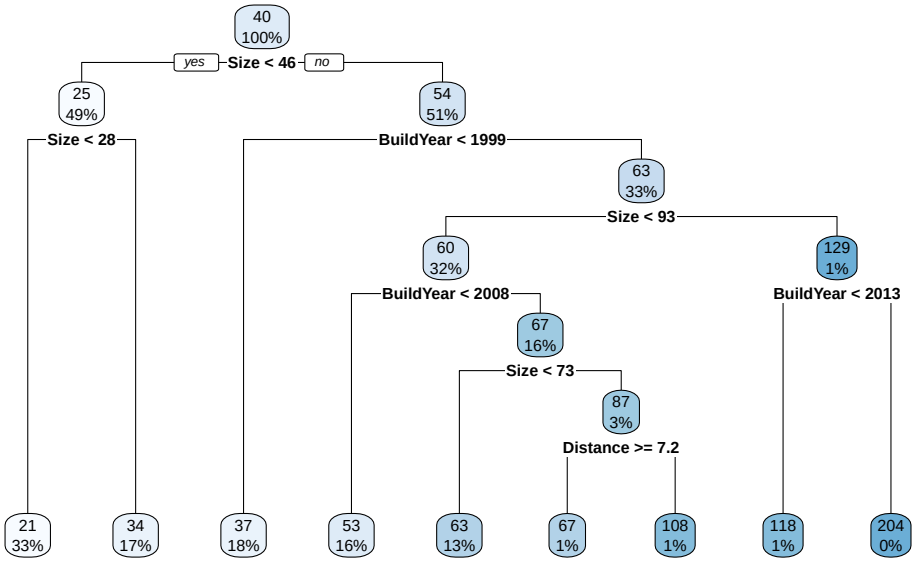
例: 交差検証



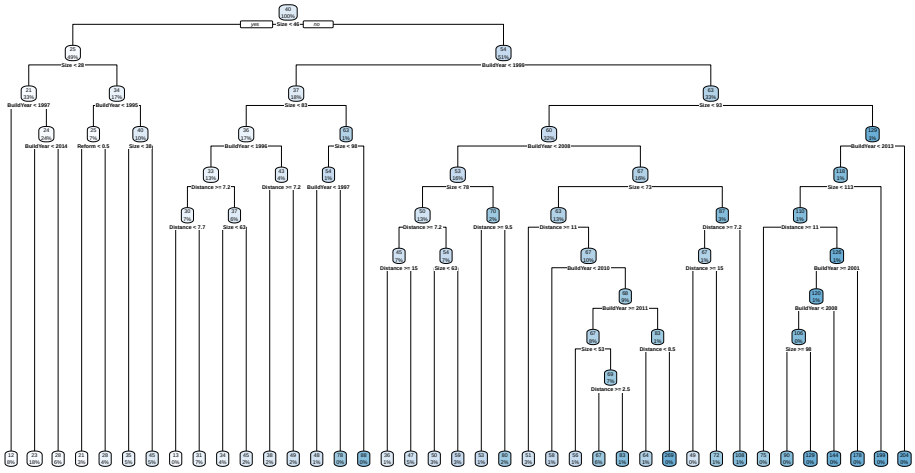
RoadMap Pruning



Example: Defaul

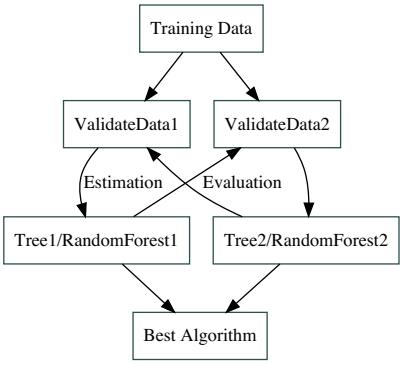


Example: Optimal



Algorithm Selection

- 交差検証はアルゴリズム選択にも応用可能



まとめ

- 交差推定を用いれば、個々の事例について、誤差項と独立した予測値を得られる
  - 因果推論への応用にも重要
- 評価への応用 (CrossValidation)
  - Algorithm の**比較** に有益