Bagging, Random Forests, Extra Trees

Introduction to Ensemble Methods

Introduction

Algorithm 1: Tree growing process

```
1  Define stopping criteria;
2  Assign training data to root node;
3  if stopping criterion is reached then
4  |  end splitting;
5  else
6  |  find the optimal split point;
7  |  split node into two subnodes at this split point;
8  |  for each node of the current tree do
9  |  |  continue tree growing process;
10 |  end
11 end
```

Ensembles

Some limitations of (single) trees

- Difficulties in modeling additive structures
- Lack of smoothness of prediction surface
- High variance / **instability** due to hierarchical splitting process

→ **Ensemble methods**

- Address instability via combining multiple prediction models
- Combine diverse models into a more robust ensemble

Ensembles

How to construct ensembles?

- Apply one method with different tuning parameter settings
- Combine models with different features
- Use one method with different subsets of the data
  - **Bagging**: Can be applied to different **base learners** (e.g. CART)
- Combine models based on different methods
  - Stacking: Build a meta-model that uses (multiple) predictions as input