2114125

Lab-8

Date / /
Page

Random Forest Ensemble

Input

→ Training data $D = \{(x_i, y_i)\}_{i=1}^{N}$, where x_i is a feature vector and y_i the target variable
→ no of trees in the forest, T
→ Number of feature to consider at each split

Output

→ a Random forest model consisting of T decision tree

Algorithm steps

1. Initialise the forest:
   - create an empty list to store T decision trees

2. For each tree $t = 1$ to T
   - Bootstrap Sampling
     - Randomly Sample N from D with replacement to create a bootstrap dataset D_t.
     Some.

- Build a decision Tree:
  - Initialize a decision tree h_t
  - At each node of tree:

Date __/__/__
Page _____

→ randomly select m features from the total M features
→ choose the best split among these m features using a criteria.
→ split the node into 2 child node based on the best split
- Continue the growth until the stopping criteria is met

3 Prediction

- For classification
  - each tree $h_t(x)$ outputs a class prediction
  - aggregate predictions from all T tree using majority voting voting to determine the final class
- For regression
  - each tree $h_t(x)$ outputs a numerical value

4 Optional: Evaluate model:

- use out of bag sample to estimate generalisation error