Decision Tree

Xuzhe Xia

October 22, 2022

0 Definitions

Decision tree is:

- A nested sequence of "if-else" decision based on the features.

- A class as a return value at the end of each sequence.

Decision stump is:

- A simple decision tree with 1 splitting rule based on threshold 1 feature.

- To learn a decision stump we need to know: which feature use to split; what value to used; classes assigned to leaves.

Notations:

- Feature matrix X: $n \times d$ dimension. (n: number of samples, d: number of features).

- Label vector y: each entry is a label of corresponds sample.

1 Accuracy-based decision tree

1.1 Decision Stump

Choice of decision stump:

- Choose the best stump based on accuracy.

```
for feature "j":
    for each threshold "t":
        set "y-yes" to most common label of objects satisfying rule.
        set "y-no" to most common label of objects not satisfying rule.
        set "y-hat" to most common label of objects "i" satisfying rule.
        compute error = number of objects where y != y-hat.
```
Listing 1: Decision stmp learning pseudo-code

- complexity: