Decision Trees 3
Friday, December 2, 2022 8:00 AM

[Sandwich joke]

Agenda

1. Splitting Numerical Features (?)
2. Underfitting & Overfitting
3. Hyper-parameters
4. Numerical version of DT
   - Regression
   - Multi-output data
   - Categorical Features
   - Fast, Space Complexity
   - Non-linear, Unstable
   - Feature importance

Numerical Features

Categorical: x_1, x_2, x_3

$I_G(f_1)$

f_1 numerical feature e.g. (Age)

1. $[x_1 < 0.2] \rightarrow I_G^{(1)}$
2. $[x_2 < 3.1] \rightarrow I_G^{(2)}$

3. $[x_1 < 5.3] \rightarrow I_G^{(3)}$

$f_1 \rightarrow I_G^{(1)}$
$f_2 \rightarrow I_G^{(1)}, I_G^{(2)}, \ldots, I_G^{(n)}$

Choose attributes split point for each f_i.

Idea: Feature Binning

$[0, 1.5), [1.5, 3.1), [3.1, 5.3), [5.3, 7.2), \ldots$

Advantages

Underfit & Overfit

Underfit

Overfit

Q: Why does a DT overfit? What is its bias-variance

Pattern has irreducibility

Decision Stump

Geometric Representation (Decision Boundary)

"Axis-Parallel Boundary"

Impact of outliers

Regular Setting

DT → Categorical

DT → Regression

DT → Multi-class classification?

Interpretability?

Can a Feature be used more than once?

Feature Importance.

Depth