Machine Learning Project 1

Fernando García Sanz - Federico Taschin

KTH Royal Institute of Technology

16th September 2019

The different decision trees, ordered by descendant difficulty, are:

- MONK-2: Check all values in the domain for each variable
- MONK-1: Check either value for a_5 or check if value of $a_1 = 1$ and $a_2 = 1$ too, $a_1 = 2$ and $a_2 = 2$, and so on
- MONK-3: Check specific values for some variables: a_5 and a_4 equal to one or $a_5 \neq 4$ and $a_2 \neq 3$

Entropy: measure of unpredictability

$$Entropy = \sum_i p(i) \log \frac{1}{p(i)}$$

Dataset	Entropy
MONK-1	1.0
MONK-2	0.957117428264771
MONK-3	0.9998061328047111

Assignment 2

Entropy for uniform and non-uniform distributions

Uniform distribution: fair die

Fair die: each number n has $P(n) = 1/6$. The entropy is:

$$E = -\sum_{i=1}^{6} P(i) \log_2 P(i) = -\log_2 \frac{1}{6} = 2.58$$

Non-uniform distribution: unfair die

Unfair die: take a die with $P(1) = 0.5$ and $P(i) = 0.1 \; \forall \; i = 2$ to 6. The entropy is:

$$E = -0.5 \log_2 0.5 - \sum_{i=2}^{6} 0.1 \log_2 0.1 = 2.16$$

Assignment 3

Information Gain 1

Information gain is the difference of entropy values of a model before and after splitting by one attribute.

Dataset	A1	A2	A3	A4	A5	A6
MONK-1	0.075	0.006	0.005	0.026	**0.287**	0.001
MONK-2	0.004	0.002	0.001	0.016	**0.017**	0.006
MONK-3	0.007	**0.294**	0.001	0.003	0.256	0.007

In this case, we select *A*5 for **MONK-1**, *A*5 for **MONK-2**, and *A*2 for **MONK-3**, since the bigger the value, the bigger the entropy reduction.

Assignment 4

Information Gain 2

$$Gain(S,A) = Entropy[S] - \sum_{k \in A} \frac{|S_k|}{|S|} Entropy[S_k]$$

Minimizing weighted average of S_k →Maximizing the gain

Maximum gain →Maximum entropy reduction in the dataset

Maximum entropy reduction →Maximum predictability

Error checking for the three decision trees:

Dataset	Error for training dataset	Error for testing dataset
MONK-1	0.0	0.171
MONK-2	0.0	0.308
MONK-3	0.0	0.056

Results matched the expectations, since the lowest error belongs to the less complex tree and the biggest error to the more complex one. Error when testing over the same training set is obviously zero.

Pruning Bias-Variance Trade-off

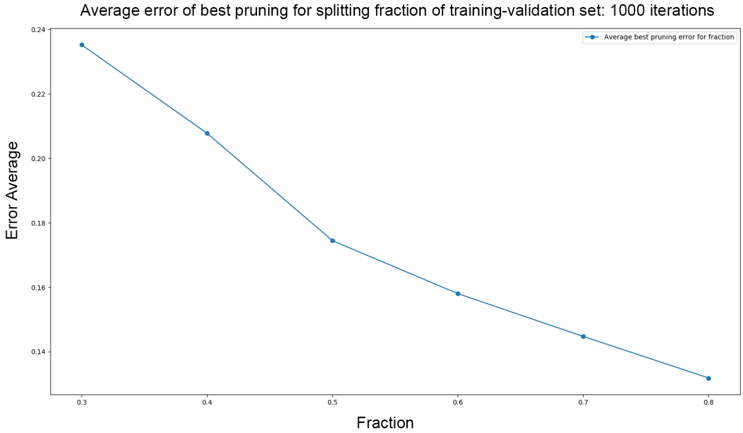
Pruning makes the tree **more general** by merging together nodes.

More general tree →Less bounded to training dataset →Less variance in predictions

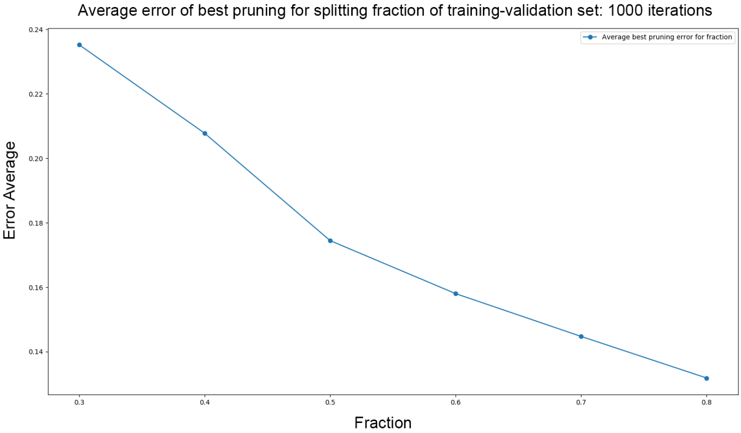
But also

More general tree →Less dataset features captured →Higher bias

MONK-1


Average error of best pruning for splitting fraction of training-validation set: 1000 iterations
Average best pruning error for fraction
Error Average
0.24
0.22
0.20
0.18
0.16
0.14
0.3
0.4
0.5
0.6
0.7
0.8
Fraction

MONK-3


Average error of best pruning for splitting fraction of training-validation set: 1000 iterations
Average best pruning error for fraction
Error Average
0.24
0.22
0.20
0.18
0.16
0.14
Fraction
0.3
0.4
0.5
0.6
0.7
0.8

MONK-1 Variances

Frequency	Variance
0.3	0.0029
0.4	0.0034
0.5	0.0041
0.6	0.0044
0.7	0.0046
0.8	0.0059

MONK-3 Variances

Frequency	Variance
0.3	0.0031
0.4	0.0016
0.5	0.0013
0.6	0.0014
0.7	0.0014
0.8	0.0024