Lab 1 – Decision Trees

Tianxiao Zhao
Feiyang Liu

The lab is about...

- Entropy and information gain
- Build up decision trees
- Reduced error pruning

Datasets: MONK 1-3, MONKTEST 1-3

Attributes: a1-a6

Assignment 0

- Hardest: Monk 2

MONK-1	$(a_1 = a_2) \vee (a_5 = 1)$
MONK-2	$a_i = 1$ for exacly two $i \in \{1, 2, \ldots, 6\}$
MONK-3	$(a_5 = 1 \wedge a_4 = 1) \vee (a_5 \neq 4 \wedge a_2 \neq 3)$

- Monk 1
  - a_3, a_4, a_6 don't have influence on the decision. Starting from a_5, the tree only needs a depth of three to decide all the outcomes.
- Monk 2
  - All attributes are independent. Therefore a depth of six is needed.
- Monk 3
  - Easiest but noise make it hard also. Still it needs a depth of three.

Assignment 1

$$Entropy(S) = -p_0 \log_2 p_0 - (1 - p_0) \log_2 (1 - p_0)$$

Entropies of 3 datasets

	Monk1	Monk2	Monk3
Entropy	1.0	0.957117	0.999806

Assignment 2

- Uniform distribution: high entropy
- Non-uniform distribution: low entropy

- Example: rolling a dice(uniform) and a fake dice(non-uniform)

$$Entropy(S) = -\sum_i p_i \log_2 p_i$$

As for normal dice $Entropy1 = 2.58$; fake dice $Entropy2 = 2.16$

Assignment 3

$$Gain(S,A) = Entropy(S) - \sum_{k \propto values(A)} \frac{S_K}{S} Entropy(S_K)$$

Information gains of 6 attributes in 3 datasets

Dataset	a1	a2	a3	a4	a5	a6
Monk 1	0.0753	0.0058	0.0047	0.0263	**0.2870**	0.0007
Monk 2	0.0038	0.0025	0.0010	0.0156	**0.0173**	0.0062
Monk 3	0.0071	**0.2937**	0.0008	0.0029	0.2560	0.0071

Bolds are used for splitting the examples at the root node

Assignment 4

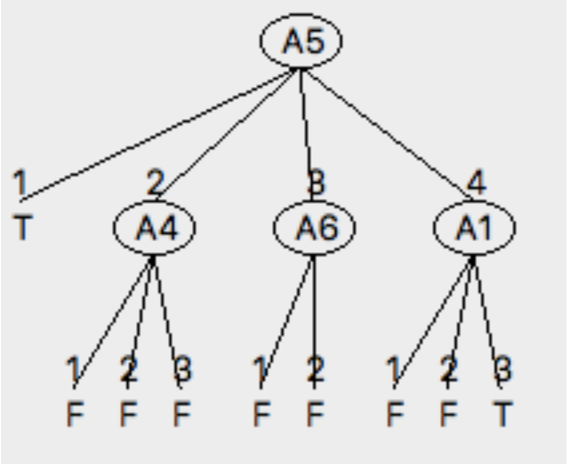
- Entropy of subset S_k: take monk 1 for example

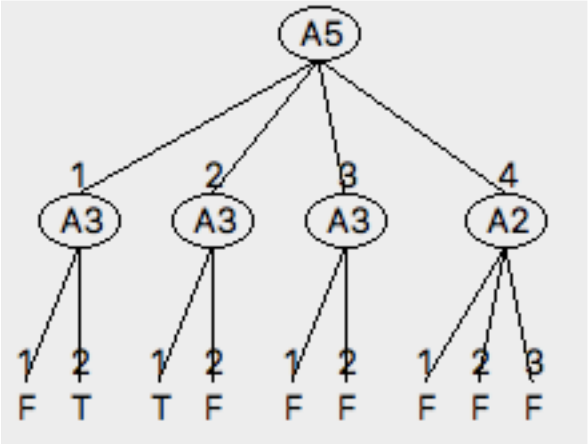
k	1	2	3	4
Entropy	0.000	0.938	0.948	0.908

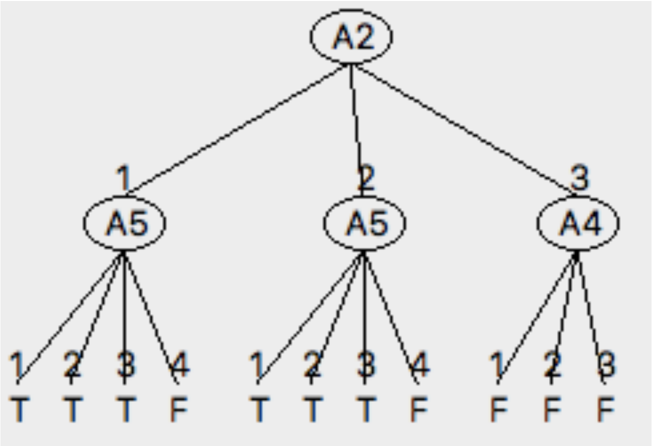
- Information gain increases <---> entropy decreases
- Smaller entropy <---> non-uniform distribution <---> dataset more concentrated <---> more certain about the classification
- Maximum information gain -> largest entropy reduction -> minimum weighted sum of entropies -> more certain about the classification

Assignment 5

- Monk 1, 2, 3 - first two levels





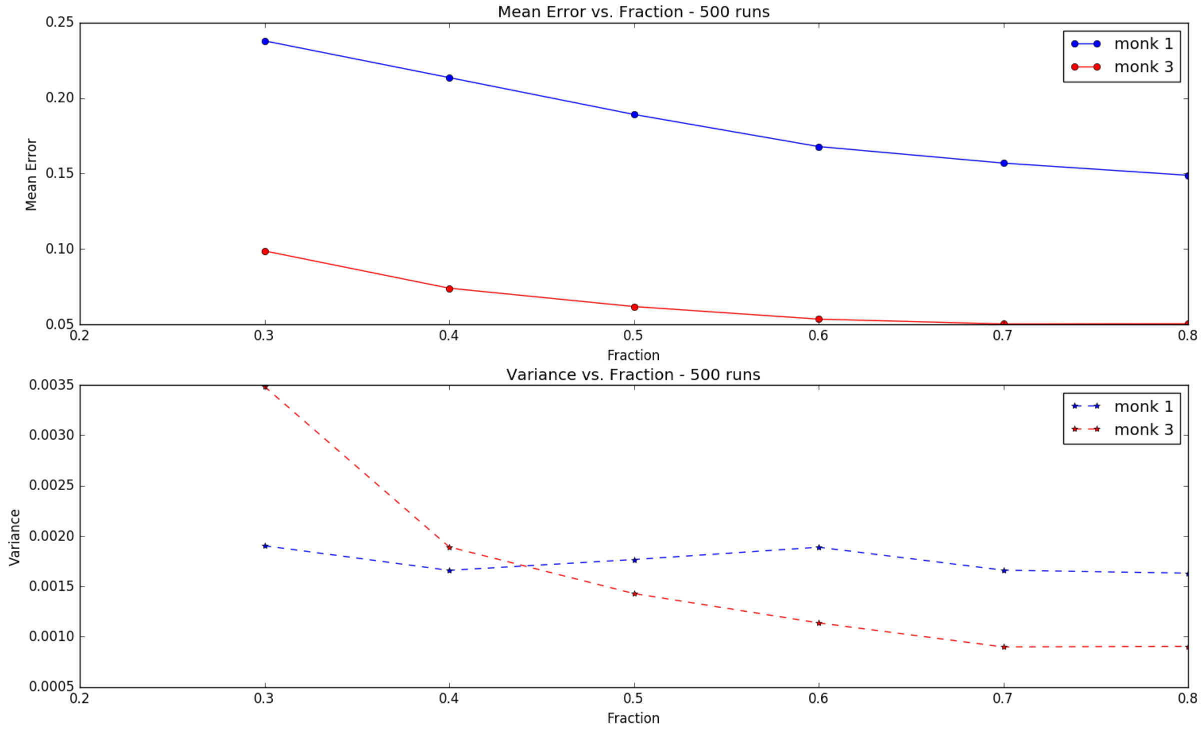


- Errors - full trees

	E_{train}	E_{test}
MONK-1	0.0	0.1713
MONK-2	0.0	0.3079
MONK-3	0.0	0.0556

Assignment 6 & 7

- Error / variance vs. fraction



fraction	Mean Error		Variance	
	Monk1	Monk3	Monk1	Monk3
0.3	0.2378	0.0987	0.00190	0.00348
0.4	0.2135	0.0741	0.00165	0.00189
0.5	0.1891	0.0618	0.00177	0.00143
0.6	0.1678	0.0536	0.00188	0.00114
0.7	0.1569	0.0504	0.00165	0.00089
0.8	0.1488	0.0505	0.00163	0.00090

- Conclusion:
  - Pruning -> less complex model (compare to original tree) -> larger bias and fewer variance
  - Larger fraction -> more training data -> more precise model -> fewer bias
  - Fraction does not directly influence model complexity