DD2431 Machine Learning - Lab 1: Decision Trees

Yamada Jun and Philipson Samuel

General Information

MONK-3 has 5% noise in a training data

MONK-1

$(a1 = a2) \vee (a5 = 1)$

MONK-2

$(a_i = 1$, for exactly two i ={1..6}

MONK-3

$((a5 = 1) \wedge (a4 = 1)) \vee ((a5 \neq 4) \wedge (a2 \neq 3))$

Assignment0

MONK-3 The comparison of the different attributes in Monk-1 just requires more splitting of the dataset, and isn't a hard thing to do. The random noice presented in Monk-3 makes it harder, though, to create a good decision tree.

Assignment1

monk-1: 1.0 monk-2: 0.957117428264771 monk-3: 0.9998061328047111

Assignment2

- Uniform Distribution

A uniform distribution makes the entropy maximize, because corresponding vriables have the constant probability and it's hard to predict an event. Suppose that we have $P(X = x_n) = \frac{1}{N}$ where X takes the value $X = [x_1, x2, ..., x_N]$(N={1. . . ..}). Then, the entropy is

$$Entropy(S) = -\Sigma_{n=1}^{N} P(X = x_n) log_2 P(x = x_n) = -\Sigma_{n=1}^{N} \frac{1}{N} log_2 \frac{1}{N} = N \times \frac{1}{N} log_2 N = log_2 N$$

- Non-Uniform Distribution The entropy of Non-Uniform Distribution gets smaler than the entropy of uniform distribution because some events happen more frequently than the other events, and it gets easier to predict what events more likely happen.
- high and low entropy distribution Suppose that we have a normal distribution, the form of entropy is $\frac{1}{2}ln(2\sigma^2\pi e)$. When $\sigma = 5$, the entropy gets large, because the shape of the distribution become flat and it is difficult to predict what events more likely happen. On the other hands, when $\sigma = 0.2$, the entropy gets low, because events around the center of distribution happen more frequently than the others.

Assignment 3

Expected Information gain

- Monk1

0.07527255560831925

0.005838429962909286

0.00470756661729721

0.02631169650768228

0.28703074971578435

0.0007578557158638421

- Monk2

0.0037561773775118823

0.0024584986660830532

0.0010561477158920196

0.015664247292643818

0.01727717693791797

0.006247622236881467

- Monk3

0.007120868396071844

0.29373617350838865

0.0008311140445336207

0.002891817288654397

0.25591172461972755

0.007077026074097326

Assignment4

The entropy implies the measure of uncertainty or unpredictability, i.e the larger the entropy is, the more uncertain an attribute had. So the maximized infomation gain is derived from the most certain attribute(the largest entropy). Then, to create an efficient decision tree we select the attribute that result in the most pure subset.

Assignment5

Monk-1

train: 1.0

test: 0.8287037037037037

Monk-2

train: 1.0

test: 0.6921296296296297

Monk-3

train: 1.0

test: 0.9444444444444444

We assume that the result of Monk-3 is not correct, since the result is the most accurate regardless having 5% additional noise. (we had better think about this result later)