**Exercise11:**

The goal is to decide if someone buys a computer or not. Derive the best decision tree by calculating a little by hand (Shannon). At least the first split.

**Calculating the Entropy and Gain for Decision Tree**

Buy Computer

|  |  |  |
| --- | --- | --- |
| Buy Computer |  |  |
| yes | no | sum |
| 12 | 8 | 20 |

Age

|  |  |  |  |
| --- | --- | --- | --- |
|  | Buy computer | Buy computer |  |
|  | yes | no | sum |
| <30 | 2 | 6 | 8 |
| 31…40 | 6 | 0 | 6 |
| >40 | 4 | 2 | 6 |

Income

|  |  |  |  |
| --- | --- | --- | --- |
|  | Buy computer | Buy computer |  |
|  | yes | no | sum |
| high | 3 | 2 | 5 |
| medium | 5 | 3 | 8 |
| low | 4 | 3 | 7 |

Student

|  |  |  |  |
| --- | --- | --- | --- |
|  | Buy computer | Buy computer |  |
|  | yes | no | sum |
| yes | 8 | 1 | 9 |
| no | 4 | 7 | 11 |

Credit rating

|  |  |  |  |
| --- | --- | --- | --- |
|  | Buy computer | Buy computer |  |
|  | yes | no | sum |
| Fair | 7 | 3 | 10 |
| Excellent | 5 | 5 | 10 |

Calculated with ID3 method from <http://www.saedsayad.com/decision_tree.htm>

Entropy Buy Computer

E(BuyComputer) = E(12,8)

= 0.971

**Entropy(BuyComputer, Age)**

E(BuyCompter, Age) = P(<30)\*E(2,6) + P(31..40)\*E(6,0) + P(>40)\*E(4,2)

= (8/20)\*0.811 + (6/20)\*0 + (6/20)\*0.918

= 0.6

**Entropy(BuyComputer, Income)**

E(BuyComputer, Income) = P(high)\*E(3,2) + P(medium)\* E(5,3) + P(low)\*E(4,3)

= (5/20)\*E(3,2) + (8/20)\*E(5,3) + (7/20)\*E(4,3)

= 0.25\*0.971 + 0.4\*0.954 + 0.35\*0.9855

= 0.96

**Entropy(BuyComputer, Student)**

E(BuyComputer, Student) = P(IsStudent)\*E(8,1)+P(noStudent)\*E(4,7)

= (9/20)\*(E8,1) + (11/20)\*E(4,7)

= 0.45\*0.5044 + 0.55\*0.9457

= 0.747

**Entropy(BuyComputer, CreditRating)**

E(BuyComputer, CreditRating) = P(Fair)\*E/7,3) + P(Excellent)\*E(5,5)

= (10/20)\*E(7,3) + 10/20\*E(5,5)

= 0.5\*0.881 + 0.5\*1

= 0.941

**Calculating the GAINs**

G(BuyComputer,Age) = E(BuyComputer) – E(BuyComputer,Age)

= 0.971 - 0.6

= 0,371

G(BuyComputer,Income) = E(BuyComputer) – E(BuyComputer,Income)

= 0.971 - 0.96

= 0,011

G(BuyComputer,Student) = E(BuyComputer) – E(BuyComputer,Student)

= 0.971 - 0.747

= 0,224

G(BuyComputer,Creditrating) = E(BuyComputer) – E(BuyComputer,Creditrating)

= 0.971 -0.941

= 0,03

**Resume**

So the most important impact ist the person’s age, followed by is a student or not.

The most unimportant property in this example is the person‘s income, **a surprising result for me**.

So our decision tree should start with age, followed by property ‘is student‘, then person’s

creditrating and at least person’s income.