

Grady Laksmono
221318327
CS 454 Review Exercise

Exercise 2

Classes: mammals, reptile, fish, amphibian, bird

Total dataset: 15

- Mammals: 5
- Reptile: 3
- Fish: 3
- Amphibian: 2
- Bird: 2

$$\begin{aligned} \text{Entropy}(S) &= -[(5/15)\log_2(5/15) + (3/15)\log_2(3/15) + (3/15)\log_2(3/15) + (2/15)\log_2(2/15) + \\ & (2/15)\log_2(2/15)] \\ &= -[-0.5283 + -0.4644 + -0.4644 + -0.3876 \\ &= 1.8447 \end{aligned}$$

$$\text{Entropy}(D) = -\sum_{i=1}^m p_i \log_2(p_i)$$

$$\text{Gain}(A) = \text{Entropy}(D) - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \text{Entropy}(D_j)$$

Body Temperatures

D1: Warm-blooded: 7

- Mammals: 5
- Bird: 2

D2: Cold-blooded: 8

- Reptile: 3
- Fish: 3
- Amphibian: 2

$$\begin{aligned} \text{Entropy}(D1) &= - [(5/7)\log_2(5/7) + (2/7)\log_2(2/7)] \\ &= - [-0.3467 + -0.5164] \\ &= 0.8631 \end{aligned}$$

$$\begin{aligned} \text{Entropy}(D2) &= - [(3/8)\log_2(3/8) + (3/8)\log_2(3/8) + (2/8)\log_2(2/8)] \\ &= - [-0.5306 + -0.5306 + -0.5] \\ &= 1.5612 \end{aligned}$$

$$\begin{aligned} \text{Gain}(\text{Body Temperature}) &= 1.8447 - [(7/15)*0.8631 + (8/15)*1.5612] \\ &= 1.8447 - [0.40278 + 0.83264] \\ &= 1.8447 - 1.23542 \\ &= 0.60928 \end{aligned}$$

$$\text{SplitInfo}(\text{Body Temperature}) = -[(7/15)\log_2(7/15) + (8/15)\log_2(8/15)]$$

$$= -[-0.5131 + -0.4837]$$

$$= 0.9968$$

$$\text{GainRatio}(\text{Body Temperature}) = 0.60928 / 0.9968 = \mathbf{0.6112}$$

Skin Cover

D1: Hair: 3

- Mammals: 3

D2: Scales: 6

- Reptile: 3
- Fish: 3

D3: None: 2

- Amphibian: 2

D4: Feathers: 2

- Bird: 2

D5: Fur: 1

- Mammals: 1

D6: Quills: 1

- Mammals: 1

$$\text{Entropy}(D1) = - [(3/3)\log_2(3/3)] = 0$$

$$\text{Entropy}(D2) = - [(3/6)\log_2(3/6) + (3/6)\log_2(3/6)]$$

$$= - [-0.5 + -0.5]$$

$$= 1$$

$$\text{Entropy}(D3) = - [(2/2)\log_2(2/2)] = 0$$

$$\text{Entropy}(D4) = - [(2/2)\log_2(2/2)] = 0$$

$$\text{Entropy}(D5) = - [(1/1)\log_2(1/1)] = 0$$

$$\text{Entropy}(D6) = - [(1/1)\log_2(1/1)] = 0$$

$$\text{Gain}(\text{Skin Cover}) = 1.8447 - [(6/15)*1] = 1.4447$$

$$\text{SplitInfo}(\text{Skin Cover}) = -[(3/15)\log_2(3/15) + (6/15)\log_2(6/15) + (2/15)\log_2(2/15) + (2/15)\log_2(2/15) + (1/15)\log_2(1/15) + (1/15)\log_2(1/15)]$$

$$= -[-0.4644 + -0.5288 + -0.3876 + -0.3876 + -0.2604 + -0.2604]$$

$$= 2.2892$$

$$\text{GainRatio}(\text{Skin Cover}) = 1.4447 / 2.2892 = \mathbf{0.6311}$$

Gives Birth

D1: Yes: 6

- Mammals: 5
- Fish: 1

D2: No: 9

- Reptile: 3
- Fish: 2
- Amphibian: 2

- Bird: 2

$$\begin{aligned}\text{Entropy}(D1) &= - [(5/6)\log_2(5/6) + (1/6)\log_2(1/6)] \\ &= - [-0.2192 + -0.4308] \\ &= 0.65\end{aligned}$$

$$\begin{aligned}\text{Entropy}(D2) &= - [(3/9)\log_2(3/9) + (2/9)\log_2(2/9) + (2/9)\log_2(2/9) + (2/9)\log_2(2/9)] \\ &= -[-0.5283 + -0.4822 + -0.4822 + -0.4822] \\ &= 1.9749\end{aligned}$$

$$\begin{aligned}\text{Gain}(\text{Gives Birth}) &= 1.8447 - [(6/15)*0.65 + (9/15)*1.9749] \\ &= 1.8477 - [0.26 + 1.18494] \\ &= 0.40276\end{aligned}$$

$$\begin{aligned}\text{SplitInfo}(\text{Gives Birth}) &= -[(6/15)\log_2(6/15) + (9/15)\log_2(9/15)] \\ &= -[-0.5288 + -0.4422] \\ &= 0.971\end{aligned}$$

$$\text{GainRaio}(\text{Gives Birth}) = 0.40276 / 0.971 = \mathbf{0.4148}$$

Aquatic Creature

D1: Yes: 4

- Fish: 3
- Mammals: 1

D2: No: 7

- Mammals: 4
- Reptile: 2
- Bird: 1

D3: Semi: 4

- Amphibian: 2
- Reptile: 1
- Bird: 1

$$\begin{aligned}\text{Entropy}(D1) &= - [(3/4)\log_2(3/4) + (1/4)\log_2(1/4)] \\ &= - [-0.3113 + -0.5] \\ &= 0.8113\end{aligned}$$

$$\begin{aligned}\text{Entropy}(D2) &= - [(4/7)\log_2(4/7) + (2/7)\log_2(2/7) + (1/7)\log_2(1/7)] \\ &= - [-0.4613 + -0.5164 + -0.4011] \\ &= 1.3788\end{aligned}$$

$$\begin{aligned}\text{Entropy}(D3) &= - [(2/4)\log_2(2/4) + (1/4)\log_2(1/4) + (1/4)\log_2(1/4)] \\ &= - [-0.5 + -0.5 + -0.5] \\ &= 1.5\end{aligned}$$

$$\text{Gain}(\text{Aquatic Creature}) = 1.8447 - [(4/15)*0.8113 + (7/15)*1.3788 + (4/15)*1.5]$$

$$\begin{aligned}
&= 1.8477 - [0.2163 + 0.64344 + 0.4] \\
&= 1.8477 - 1.25974 \\
&= 0.58796
\end{aligned}$$

$$\begin{aligned}
\text{SplitInfo(Aquatic Creature)} &= -[(4/15)\log_2(4/15) + (7/15)\log_2(7/15) + (4/15)\log_2(4/15)] \\
&= -[-0.5085 + -0.5131 + -0.5085] \\
&= 1.5301
\end{aligned}$$

$$\text{GainRatio(Aquatic Creature)} = 0.58796 / 1.5301 = \mathbf{0.3843}$$

Aerial Creature

D1: Yes: 2

- Mammal: 1
- Bird: 1

D2: No: 13

- Mammal: 4
- Reptile: 3
- Fish: 3
- Amphibian: 2
- Bird: 1

$$\begin{aligned}
\text{Entropy(D1)} &= - [(1/2)\log_2(1/2) + (1/2)\log_2(1/2)] \\
&= - [-0.5 + -0.5] \\
&= 1.0
\end{aligned}$$

$$\begin{aligned}
\text{Entropy(D2)} &= - [(4/13)\log_2(4/13) + (3/14)\log_2(3/14) + (3/14)\log_2(3/14) + (2/14)\log_2(2/14) + \\
&(1/14)\log_2(1/14)] \\
&= - [-0.5232 + -0.0889 + -0.4762 + -0.4011 + -0.2720] \\
&= 1.7614
\end{aligned}$$

$$\begin{aligned}
\text{Gain(Aerial Creature)} &= 1.8447 - [(2/15)*1.0 + (13/15)*1.7614] \\
&= 1.8447 - [0.1333 + 1.5265] \\
&= 0.1849
\end{aligned}$$

$$\begin{aligned}
\text{SplitInfo(Aerial Creature)} &= - [(2/15)\log_2(2/15) + (13/15)\log_2(13/15)] \\
&= - [-0.3876 + -0.1789] \\
&= 0.5665
\end{aligned}$$

$$\text{GainRatio(Aerial Creature)} = 0.1849/0.5665 = \mathbf{0.3264}$$

Has Legs

D1: Yes: 10

- Mammal: 4
- Amphibian: 2
- Reptile: 2
- Bird: 2

D2: No: 5

- Reptile: 1
- Fish: 3
- Mammal: 1

$$\begin{aligned}\text{Entropy}(D1) &= - [(4/10)\log_2(4/10) + (2/10)\log_2(2/10) + (2/10)\log_2(2/10) + (2/10)\log_2(2/10)] \\ &= - [-0.52877 + -0.4644 + -0.4644 + -0.4644] \\ &= 1.92197\end{aligned}$$

$$\begin{aligned}\text{Entropy}(D2) &= - [(1/5)\log_2(1/5) + (3/5)\log_2(3/5) + (1/5)\log_2(1/5)] \\ &= - [-0.4644 + -0.4422 + -0.4644] \\ &= 1.371\end{aligned}$$

$$\begin{aligned}\text{Gain}(\text{Has Legs}) &= 1.8447 - [(10/15)*1.92197 + (5/15)*1.371] \\ &= 1.8447 - [1.2813 + 0.457] \\ &= 1.8447 - 1.7383 \\ &= 0.1064\end{aligned}$$

$$\begin{aligned}\text{SplitInfo}(\text{Has Legs}) &= - [(10/15)\log_2(10/15) + (5/15)\log_2(5/15)] \\ &= - [-0.39 + -0.5283] \\ &= 0.9183\end{aligned}$$

$$\text{GainRatio}(\text{Has Legs}) = 0.1064 / 0.9183 = \mathbf{0.1159}$$

Hibernate

D1: Yes: 5

- Reptile: 1
- Amphibian: 2
- Mammal: 2

D2: No: 10

- Mammal: 3
- Fish: 3
- Reptile: 2
- Bird: 2

$$\begin{aligned}\text{Entropy}(D1) &= - [(1/5)\log_2(1/5) + (2/5)\log_2(2/5) + (2/5)\log_2(2/5)] \\ &= - [-0.4644 + -0.5288 + -0.5288] \\ &= 1.522\end{aligned}$$

$$\begin{aligned}\text{Entropy}(D2) &= - [(3/10)\log_2(3/10) + (3/10)\log_2(3/10) + (2/10)\log_2(2/10) + (2/10)\log_2(2/10)] \\ &= - [-0.5211 + -0.5211 + -0.4644 + -0.4644] \\ &= 1.971\end{aligned}$$

$$\begin{aligned}\text{Gain}(\text{Hibernate}) &= 1.8447 - [(5/15)*1.522 + (10/15)*1.971] \\ &= 1.8447 - [0.5073 + 1.314]\end{aligned}$$

$$= 0.0234$$

$$\begin{aligned}\text{SplitInfo(Hibernate)} &= - [(5/15)\log_2(5/15) + (10/15)\log_2(10/15)] \\ &= - [-0.5283 + -0.390] \\ &= 0.9183\end{aligned}$$

$$\text{GainRatio(Hibernate)} = 0.0234 / 0.9183 = \mathbf{0.0255}$$

$$\text{GainRatio(Body Temperature)} = \mathbf{0.6112}$$

$$\text{GainRatio(Skin Cover)} = \mathbf{0.6311}$$

$$\text{GainRatio(Gives Birth)} = \mathbf{0.4148}$$

$$\text{GainRatio(Aquatic Creature)} = \mathbf{0.3843}$$

$$\text{GainRatio(Aerial Creature)} = \mathbf{0.3264}$$

$$\text{GainRatio(Has Legs)} = \mathbf{0.1159}$$

$$\text{GainRatio(Hibernate)} = \mathbf{0.0255}$$

Therefore, we split at **Skin Cover**