Lab 3

Knowledge-Based Systems in Bioinformatics

Rasmus Hammar

Task 1

A = (U,A ∪ {d})

```
info_TA <- c(
  -(4 / 9) * log2(4 / 9) - (5 / 9) * log2(5 / 9)
)
```

Info([4,5]) = 0.9910761 bits.

Table 1. Decision system A discretized using equal frequency with three levels.

Cuts	decision
[91, 105)	No
[45, 91)	{Yes, No}
[21, 45)	Yes

Calculate bits.

```
info_T1 <- c(
  -(0 / 3) * log2(0 / 3) - (3 / 3) * log2(3 / 3),
  -(1 / 3) * log2(1 / 3) - (2 / 3) * log2(2 / 3),
  -(3 / 3) * log2(3 / 3) - (0 / 3) * log2(0 / 3)
)
info_T1[1] <- 0 # defined as 0 bits
info_T1[3] <- 0 # defined as 0 bits
info_T1 <- sum(info_T1 * c(3 / 9, 3 / 9, 3 / 9))
```

Info([0,3],[1,2],[3,0]) = 0.3060986 bits.

Gain(feature_1) = info([4,5]) - info([0,3],[1,2],[3,0]) = 0.6849774 bits.

Gain_ratio(feature_1) = gain(feature_1) / split_info(feature_1)

Table 2. Decision system A discretized using naive method.

Cuts	decision
[78, 105)	No
[62, 78)	Yes
[45, 62)	No
[21, 45)	Yes

Calculate bits.

```
info_T2 <- c(
  -(0 / 4) * log2(0 / 4) - (4 / 4) * log2(4 / 4),
  -(1 / 1) * log2(1 / 1) - (0 / 1) * log2(0 / 1),
  -(0 / 1) * log2(0 / 1) - (1 / 1) * log2(1 / 1),
  -(3 / 3) * log2(3 / 3) - (0 / 3) * log2(0 / 3)
)
info_T2 <- c(0, 0, 0, 0) # defined as 0 bits
info_T2 <- sum(info_T2 * c(4 / 9, 1 / 9, 1 / 9, 3 / 9))
```

Info([0,3],[1,2],[3,0]) = 0 bits.

Gain = info([4,5]) - info([0,3],[1,2],[3,0]) = 0.9910761 bits.

B = (U,B ∪ {d})

Table 3. Decision system A discretized using equal frequency with three levels.

Cuts	decision
[0.67, 0.97)	{A, B}
[0.29, 0.67)	{A, B}
[0.04, 0.29)	{A, B}

Table 4. Decision system A discretized using naive method.

Cuts	decision
[0.89, 0.97)	A
[0.75, 0.89)	B
[0.39, 0.75)	A
[0.17, 0.39)	B
[0.08, 0.17)	A
[0.03, 0.08)	B

Task 2