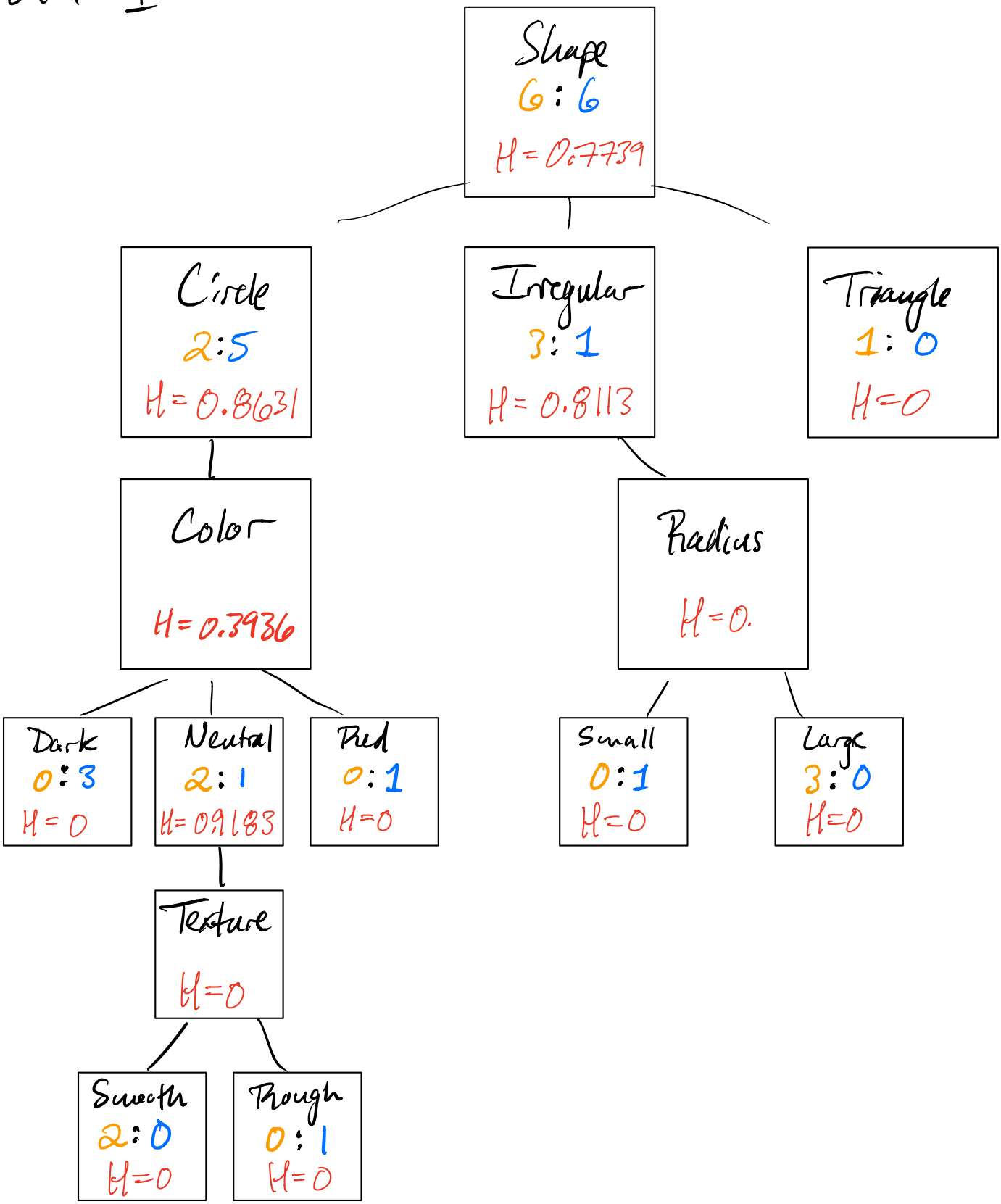
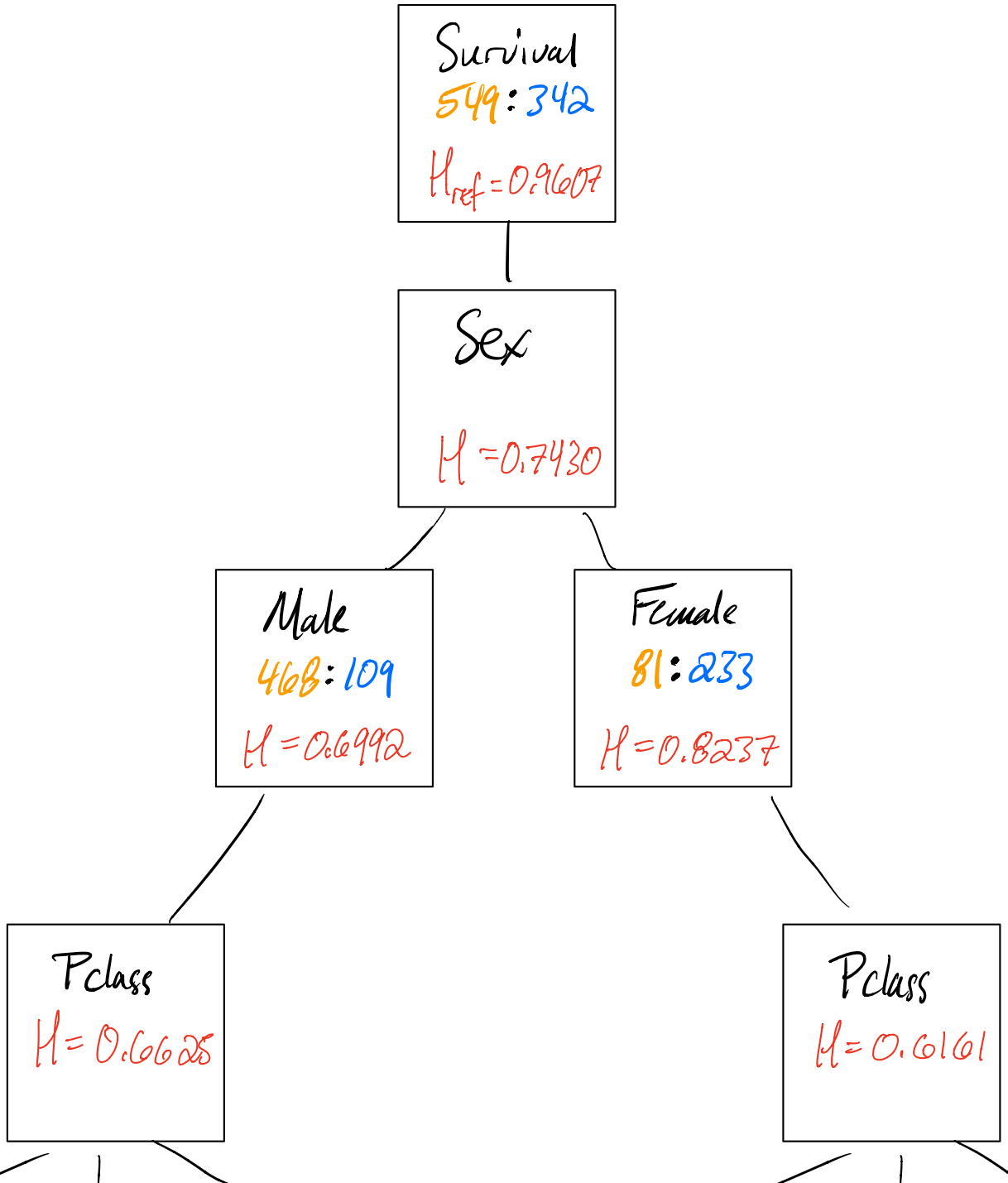
Final #1

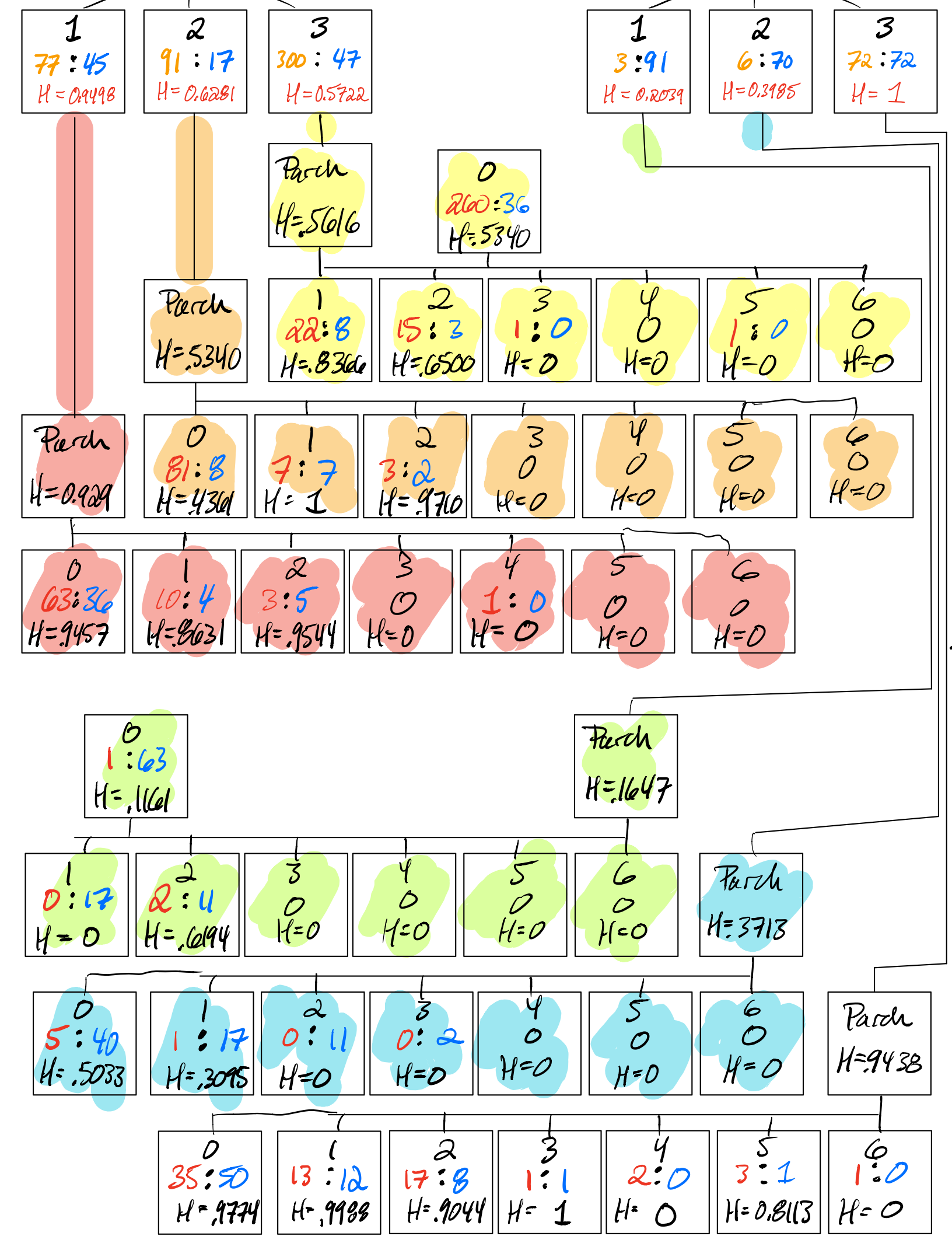
Part 1

Liam Jackson
BU U40546227



Part 2:


Survival
549 : 342
$H_{ref} = 0.9607$
Sex
$H = 0.7430$
Male
468 : 109
$H = 0.6992$
Female
81 : 233
$H = 0.8237$
Pclass
$H = 0.6625$
Pclass
$H = 0.6161$


1
77 : 45
H = 0.9498
2
91 : 17
H = 0.6281
3
300 : 47
H = 0.5722
1
3 : 91
H = 0.2039
2
6 : 70
H = 0.3985
3
72 : 72
H = 1
Parch
H=.5616
0
260 : 36
H=.5340
Parch
H=.5340
1
22 : 8
H=.8366
2
15 : 3
H=.6500
3
1 : 0
H=0
4
0
H=0
5
1 : 0
H=0
6
0
H=0
Parch
H=0.929
0
81 : 8
H=.4361
1
7 : 7
H=1
2
3 : 2
H=.9710
3
0
H=0
4
0
H=0
5
0
H=0
6
0
H=0
0
63 : 36
H=.9457
1
10 : 4
H=.8631
2
3 : 5
H=.9544
3
0
H=0
4
1 : 0
H=0
5
0
H=0
6
0
H=0
0
1 : 63
H=.1161
Parch
H=.1647
1
0 : 17
H=0
2
2 : 4
H=.6194
3
0
H=0
4
0
H=0
5
0
H=0
6
0
H=0
Parch
H=.3718
0
5 : 40
H=.5033
1
1 : 17
H=.3095
2
0 : 11
H=0
3
0 : 2
H=0
4
0
H=0
5
0
H=0
6
0
H=0
Parch
H=.9438
0
35 : 50
H=.9774
1
13 : 12
H=.9988
2
17 : 8
H=.9044
3
1 : 1
H=1
4
2 : 0
H=0
5
3 : 1
H=0.8113
6
1 : 0
H=0

$$H = -\sum_{i=1}^{\#classes} p_i \log_2(p_i)$$

A) Your dataset: 12 biopsies Total ($N = 12$)

Instance	$\vec{x}_1$ Shape	$\vec{x}_2$ Radius	$\vec{x}_3$ Concavity	$\vec{x}_4$ Texture	$\vec{x}_5$ Color	Class label (malignant or not?)
1	Circle	L	convex	Smooth	Dark	+ 1
2		L	concave		Dark	+ 1
3		L	flat		Red	+ 1
4	Irregular	S	concave	Rough	Dark	+ 1
5	Circle	L	flat		Neutral	+ 1
6		L	Concave		Dark	+ 1
7		L	convex	Smooth	Neutral	− 0
8	Irregular	L	concave		Red	− 0
9	Triangle	S	convex	Rough	Dark	− 0
10		L	flat	smooth	Neutral	− 0
11	Irregular	L	concave		Dark	− 0
12	Irregular	L	concave		Red	− 0

malignant: 1–6; Benign: 7–12

Table 1: The tissue biopsy example that we did in class !

$$H_{ref} = -\frac{6}{12}\log_2\left(\frac{6}{12}\right) - \frac{6}{12}\log_2\left(\frac{6}{12}\right) = 1.0$$

$P_0 = \frac{6}{12}$ $P_1 = \frac{6}{12}$

BTW, I wrote a MATLAB Fxn to calc these, hence not much math shown...

Shape

Shannons:

Circle: $-\frac{2}{7}\log_2\left(\frac{2}{7}\right) - \frac{5}{7}\log_2\left(\frac{5}{7}\right)$ $P_0 = \frac{2}{7}$ $P_1 = \frac{5}{7}$

Irreg: $-\frac{3}{4}\log_2\left(\frac{3}{4}\right) - \frac{1}{4}\log_2\left(\frac{1}{4}\right)$ $P_0 = \frac{3}{4}$ $P_1 = \frac{1}{4}$

Triangle: $-\log_2(1) - 0\log_2(0)$ $P_0 = \frac{1}{1}$ $P_1 = \frac{0}{1}$

$$H_{shape} = \frac{7}{12}(0.8631) + \frac{4}{12}(0.8113) + \frac{1}{12}(0)$$

$$H_{shape} = 0.7739 \Rightarrow IG_{ref \to shape} = 0.2261$$

Radius

Shannons:

Small: $-\frac{1}{2}\log_2\left(\frac{1}{2}\right) - \frac{1}{2}\log_2\left(\frac{1}{2}\right)$ $P_0 = \frac{1}{2}$ $P_1 = \frac{1}{2}$

Large: $-\frac{5}{10}\log_2\left(\frac{5}{10}\right) - \frac{5}{10}\log_2\left(\frac{5}{10}\right)$ $P_0 = \frac{5}{10}$ $P_1 = \frac{5}{10}$

$$H_{Radius} = \frac{2}{12}(1) + \frac{10}{12}(1)$$

$$H_{Radius} = 1 \Rightarrow IG_{ref \to Radius} = 0$$

Concavity

Shannons: Concave: $-\frac{3}{6}\log_2\left(\frac{3}{6}\right)-\frac{3}{6}\log_2\left(\frac{3}{6}\right)$ $p_0=3/6$ $p_1=3/6$; Convex: $-\frac{2}{3}\log_2\left(\frac{2}{3}\right)-\frac{1}{3}\log_2\left(\frac{1}{3}\right)$ $p_0=2/3$ $p_1=1/3$; Flat: $-\frac{1}{3}\log_2\left(\frac{1}{3}\right)-\frac{2}{3}\log_2\left(\frac{2}{3}\right)$ $p_0=1/3$ $p_1=2/3$

$$H_{conc} = \frac{6}{12}(1) + \frac{3}{12}(0.9183) + \frac{3}{12}(0.9183)$$

$$H_{conc} = 0.9591 \Rightarrow IG_{ref\to conc} = 0.0409$$

Texture

Rough: $-\frac{1}{4}\log_2\left(\frac{1}{4}\right)-\frac{3}{4}\log_2\left(\frac{3}{4}\right)$ $p_0=1/4$ $p_1=3/4$

Smooth: $-\frac{5}{8}\log_2\left(\frac{5}{8}\right)-\frac{3}{8}\log_2\left(\frac{3}{8}\right)$ $p_0=5/8$ $p_1=3/8$

$$H_{tex} = \frac{4}{12}(0.8113) + \frac{8}{12}(0.9544)$$

$$H_{tex} = 0.9067 \Rightarrow IG_{ref\to tex} = 0.0933$$

Color

Dark: $-\frac{2}{6}\log_2\left(\frac{2}{6}\right)-\frac{4}{6}\log_2\left(\frac{4}{6}\right)$ $p_0=2/6$ $p_1=4/6$

Neutral: $-\frac{2}{3}\log_2\left(\frac{2}{3}\right)-\frac{1}{3}\log_2\left(\frac{1}{3}\right)$ $p_0=2/3$ $p_1=1/3$

Red: $-\frac{2}{3}\log_2\left(\frac{2}{3}\right)-\frac{1}{3}\log_2\left(\frac{1}{3}\right)$ $p_0=2/3$ $p_1=1/3$

$$H_{color} = \frac{6}{12}(0.9183) + \frac{3}{12}(0.9183) + \frac{3}{12}(0.9183)$$

$$H_{color} = 0.9183 \Rightarrow IG_{ref\to color} = 0.0817$$

$IG_{ref\to shape} = 0.2261$ is max $\Rightarrow$ Shape first decision