Oliver De Sa
1005457943

LMP1210 Assignment 2

Answer Locations:

1. The hand written calculations are attached to this PDF, including the final interpretation at the end.
2. Both the code and the written answers for question 2 can be found at: https://colab.research.google.com/drive/1ShfHdZBjUI3rl0x2kK-oiM3EiveLEPM0?usp=sharing
3. Both the code and the written answers for question 3 can be found at: https://colab.research.google.com/drive/1ShfHdZBjUI3rl0x2kK-oiM3EiveLEPM0?usp=sharing
4. Code for question 4 can be found at: https://colab.research.google.com/drive/1_b9zzellzp1Z-MRQyHqapsc-19MeZBsp?usp=sharing
5. The hand written calculations, along with computational graph are attached to this PDF.
6. Both the code and the written answers for question 6 can be found at: https://colab.research.google.com/drive/1ShfHdZBjUI3rl0x2kK-oiM3EiveLEPM0?usp=sharing

1. Root Entropy

$$H(Y) = -\frac{4}{6}\cdot\log_2\left(\frac{4}{6}\right) - \frac{2}{6}\cdot\log_2\left(\frac{2}{6}\right)$$

$$\approx 0.918$$

Heart Failure?

Chest Pain	Yes	No
Yes	4	0
No	0	2

$$IG(\text{chest pain}) = 0.918 - H(Y|CP)$$

$$H(Y|CP) = \left(\frac{4}{6}\left(-\frac{4}{4}\cdot\log_2\frac{4}{4} - \frac{0}{4}\log_2\frac{0}{4}\right) + \left(\frac{2}{6}\left(-\frac{2}{2}\cdot\log_2\frac{2}{2} - \frac{0}{2}\cdot\log_2\frac{0}{2}\right)\right)\right.$$

$$= 0$$

$$IG(CP) = 0.918$$

Heart Failure?

Male	Yes	No
Yes	3	2
No	1	0

$$IG(\text{Male}) = 0.918 - H(Y|M)$$

$$H(Y|M) = \left(\frac{5}{6}\left(-\frac{3}{5}\cdot\log_2\frac{3}{5} - \frac{2}{5}\log_2\frac{2}{5}\right)\right) + \left(\frac{1}{6}\left(-1\log_2 1 - 0\log_2 0\right)\right)$$

$$\approx 0.809 + 0$$

$$= 0.809$$

$$IG(M) = 0.918 - 0.809$$

$$\approx 0.108$$

Smokes	Heart failure? Yes	No
Yes	2	1
NO	2	1

$IG(\text{Smokes}) = 0.918 - H(Y|S)$

$H(Y|S) = \left(\frac{3}{6}\left(-\frac{2}{3}\cdot\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right)\right) + \left(\frac{3}{6}\left(-\frac{2}{3}\log_2\frac{2}{3} - \frac{1}{3}\log_2\frac{1}{3}\right)\right)$

$\doteq 0.4591 + 0.4591$

$= 0.918$

$IG(S) = 0.918 - 0.918$

$\doteq 0$

Exercise	Heart failure? Yes	No
Yes	2	2
NO	2	0

$IG(\text{Exercise}) = 0.918 - H(Y|E)$

$H(Y|E) = \left(\frac{4}{6}\left(-\frac{1}{2}\cdot\log_2\frac{1}{2} - \frac{1}{2}\log_2\frac{1}{2}\right)\right) + \left(\frac{2}{6}\left(-1\log_2 1 - 0\log_2 0\right)\right)$

$\doteq 0.666 + 0$

$= 0.666$

$IG(E) = 0.918 - 0.666$

$\doteq 0.252$

Therefore, chest pain has the largest information gain at the root level of the decision tree. It is the strongest predictor of whether or not a patient will have heart failure.

$f(x, y, w) = (x-y)^2 \cdot e^w$

Baseline: $\bar{f} = 1$

$\bar{z} = 2(x-y)$

$\bar{x} = 2(x-y) \cdot 2$

$\bar{y} = 2(x-y) \cdot (-2)$

$\bar{w} = e^w$