Homework 4 – Due 4/27 at 9 AM Eastern Time – Prepared by Michael Wacey

1. Three persons A, B and C have applied for a job in a private company. The chance of their selections is in the ratio 1 : 4 : 2. The probabilities that A, B and C can introduce changes to improve the profits of the company are 0.4, 0.5 and 0.6, respectively. If the change does not take place, find the probability that it is due to the appointment of C.

1. **Selection Probabilities:**
   * A: 1/7 (1 chance out of the sum of ratios 1 + 4 + 2)
   * B: 4/7
   * C: 2/7
2. **Probability of Change and No Change:**
   * Probability of A and Change 1/7 \* 4/10 = 4/70
   * Probability of A and No Change 1/7 \* 6/10 = 6/70
   * Probability of B and Change 4/7 \* 5/10 = 20/70
   * Probability of B and No Change 4/7 \* 5/10 = 20/70
   * Probability of C and Change 2/7 \* 6/10 = 12/70
   * Probability of C and No Change 2/7 \* 4/10 = 8/70
   * Probability of Change = 4/70 + 20/70 + 12/70 = 36/70
   * Probability of No Change = 6/70 + 20/70 + 8/70 = 34/70
   * Note: I did this on a tree on paper, but it was easier to write statements in Word.
3. **Probability of No Change given the Selection of C:**
   * P(NC|C) – The probability of no change given the selection of C
   * Conditional Probability – P(A|B) = P(A and B) / P(B)
   * P(NC | C) = P(NC and C) / P(C) = (8/70) / (2/7) = (8/70) \* (7/2) = 56 / 140 = 2/5

2. Consider a statistical experiment where we model a game consisting of drawing two types of coins from a bag (with replacement) for a total of four possible outcomes. The state space or sample space Ω of this experiment is then ($, $), ($, ￡), (￡, $), (￡, ￡). Let us assume that the composition of the bag of coins is such that a draw returns at random a $ with probability 0.4. What is the probability that a $ is drawn one time?

($,$) = 2

($,£) = 1

(£,$) = 1

(£,£) = 0

P(X=2) = P($,$) = P($) \* P($) = 0.4 \* 0.4 = 0.16

P(X=1) = P($,£) + P(£,$) = (0.4 \* 1 – 0.4) + (1- - 0.4 \* 0.4) = 0.48

P(X=1) = P(£,£) = P(£) \* P(£) = (1 – 0.4) \* (1 – 0.4) = 0.36

Since 0.16 + 0.48 + 0.36 = 1, all conditions are accounted for, so the probability that a $ is drawn one time is 0.48.

3.

A white background with black text

Description automatically generated

a. What is the probability of if P(A) = 0.1, P(B) is 0.2 and the probability of ?

**If P(A U B)’ = 0.75, then the P(A U B) = 0.25.**

**Given that P(A U B) = P(A) + P(B) – P(A∩B), then P(A∩B) = P(A) + P(B) - P(A U B)**

**P(A∩B) = 0.1 + 0.2 – 0.25 = 0.3 – 0.25 = 0.05**

b. What is P(A|B)?

**P(A|B) = P(A∩B) / P(B) = 0.05 / 0.2 = 0.25**

c. What is P(B|A)?

**P(B|A) = P(B∩A) / P(A) = P(A∩B) / P(A) = 0.05 / 0.1 = 0.5**

4. Four people are tested for presence of antibodies for a particular disease, and we expect the antibodies to have a 1/10 probability of occurrence in each person. What is the probability of at least three individuals being tested before one of them is found to have the antibodies?

**Did not do per your email.**

5. Given the following dataset, calculate the following if the Target Variable is the Temperature:

A table with different types of temperature

Description automatically generated

1. Entropy for the Target Variable

Temperature = {Hot, Mild, Cool}

E(Temperature) = Σi=1 -pilog2pi

= -((4/14)log2(4/14) + (6/14)log2(6/14) + (4/14)log2(4/14))

= -(0.286 log2 0.286 + 0.429 log2 0.429 + 0.286 log2 0.286)

= -(-0.516 + -0.524 + -0.516)

= **1.557**

1. Entropy for both attributes

Outlook = {Sunny, Overcast, Rainy)

E(Temperature | Outlook = Sunny) = Σi=1 -pilog2pi

= -((2/5)log2(2/5) + (2/5)log2(2/5) + (1/5)log2(1/5))

= -(0.4 log2 0.4 + 0.4 log2 0.4 + 0.2 log2 0.2)

= -(-0.529 + -0.529 + -0.464)

= **1.522**

E(Temperature | Outlook = Overcast) = Σi=1 -pilog2pi

= -((2/4)log2(2/4) + (1/4)log2(1/4) + (1/4)log2(1/4))

= -(0.5 log2 0.5 + 0.25 log2 0.25 + 0.25 log2 0.25)

= -(-0.5 + -0.5 + -0.5)

= **1.5**

E(Temperature | Outlook = Rainy) = Σi=1 -pilog2pi

= -( (3/5)log2(3/5) + (2/5)log2(2/5))

= -(0.6 log2 0.6 + 0.4 log2 0.4)

= -(-0.442 + -0.529)

= **0.971**

Humidity = {High, Normal)

E(Temperature | Humidity = High) = Σi=1 -pilog2pi

= -((3/7)log2(3/7) + (4/7)log2(4/7))

= -(0.429 log2 0.429 + 0.571 log2 0.571)

= -(-0.524 + -0.461)

**= 0.985**

E(Temperature | Humidity = Normal) = Σi=1 -pilog2pi

= -((1/7)log2(1/7) + (2/7)log2(2/7) + (4/7)log2(4/7))

= -(0.143 log2 0.143 + 0.286 log2 0.286 + 0.571 log2 0.571)

= -(-0.401 + -0.516 + -0.461)

= **1.379**

1. Information Gain for both Attributes

IG(Outlook) = E(Temperature) – (P(Outlook=Sunny)\*E(Temperature | Outlook=Sunny)

+P(Outlook=Overcast)\*E(Temperature | Outlook=Overcast)

+P(Outlook=Rainy)\*E(Temperature | Outlook=Rainy)

)

= 1.557 – ((5/14 \* 1.522) + (4/14 \* 1.5) + (5/14 \* 0.971))

= **0.238**

IG(Humidity) = E(Temperature) – (P(Humidity = High)\*E(Temperature | Normal = High)

+P(Humidity = Normal)\*E(Temperature | Normal = Normal)

)

= 1.557 – ((7/14 \* 0.985) + (7/14 \* 1.379))

= **0.375**

1. Should outlook or humidity be the first node in decision tree?

Since the Humidity has a higher Information Gain, it should be the first node in the decision tree.