Fullname: Phạm Văn Phúc

Student’s ID: 522H0068

**Homework 3**

**Question 1. Classification with Decision tree and Bayes**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Age | Income | Student | Credit | Target Buys Computer |
| 1 | Youth | High | No | Fair | No |
| 2 | Youth | High | No | Excellent | No |
| 3 | Middle | High | No | Fair | Yes |
| 4 | Senior | Medium | No | Fair | Yes |
| 5 | Senior | Low | Yes | Fair | Yes |
| 6 | Senior | Low | Yes | Excellent | No |
| 7 | Middle | Low | Yes | Excellent | Yes |
| 8 | Youth | Medium | No | Fair | No |
| 9 | Youth | Low | Yes | Fair | Yes |
| 10 | Senior | Medium | Yes | Fair | Yes |
| 11 | Youth | Medium | Yes | Excellent | Yes |
| 12 | Middle | Medium | No | Excellent | Yes |
| 13 | Middle | High | Yes | Fair | Yes |
| 14 | Senior | Medium | No | Excellent | No |

Which class the following object is belonging to?

Age=youth, Income=medium, Student=yes, Credit\_rating=fair

**My answer:**

* **Decision Tree**

***Calculate Total Entropy:***

Total Samples: 14  
Samples **Yes**: 9  
Samples **No**: 5

Entrophy: 

***Calculate Information Gain for each attribute:***

* + Gain (Age):

**Age** has 3 values: Youth, Middle, Senior

* + - Age = Youth (sample 1,2,8,9,11):

5 samples: 2 Yes, 3 No



* + - Age = Middle (sample 3,7,12,13):

4 samples: 4 Yes, 0 No → Entropy = 0

* + - Age = Senior (sample 4,5,6,10,14):

5 samples: 3 Yes, 2 No → Entropy ≈ 0.971



Information Gain for Age:



* + Gain (Income)

**Income** has 3 values: High, Medium, Low

* + - High (sample 1,2,3,13):

4 samples: 2 Yes, 2 No



* + - Medium (sample 4,8,10,11,12,14):

6 samples: 4 Yes, 2 No



* + - Low (sample 5,6,7,9):

4 samples: 3 Yes, 1 No



Information Gain cho Income:



* + Gain (Student)

**Student** has 2 values: Yes, No

* + - Student = Yes (sample 5,6,7,9,10,11,13):

7 samples: 6 Yes, 1 No



* + - Student = No (sample 1,2,3,4,8,12,14):

7 samples: 3 Yes, 4 No



Information Gain for Student:



* + Gain (Credit)

**Credit** has 2 values: Fair, Excellent

* + - Fair (sample 1,3,4,5,8,9,10,13):

8 sample: 6 Yes, 2 No



* + - Excellent (sample 2,6,7,11,12,14):

6 sample: 3 Yes, 3 No



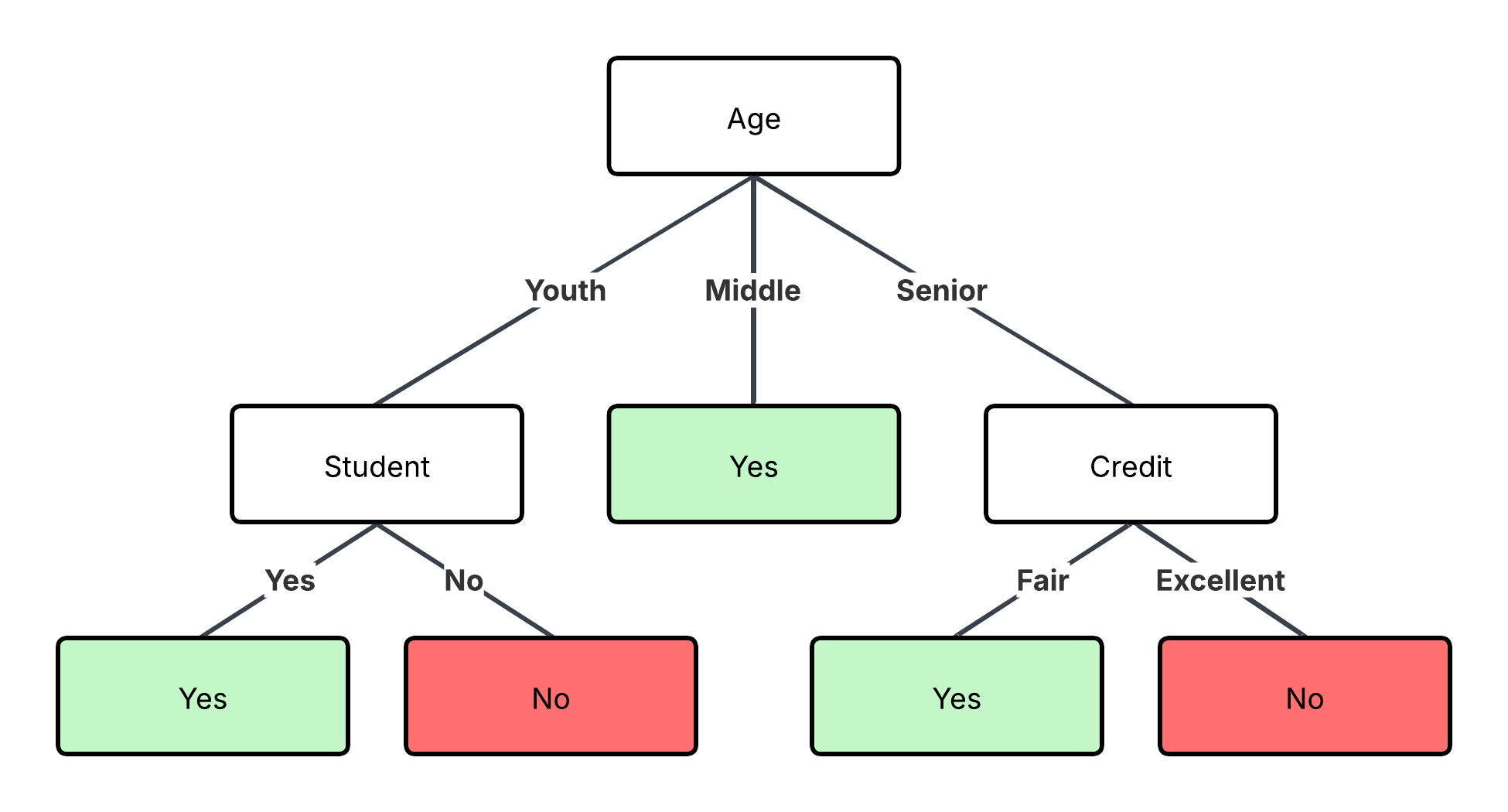
Information Gain for Credit:



Have:

* + Gain (Age) = 0.246
  + Gain (Income) = 0.029
  + Gain (Student) = 0.151
  + Gain (Credit) = 0.048

So Gain (0.246) is the largest so it will decide Age as the root of the Decision Tree

***Decision Tree: ***

Since Income only reduces Entropy very slightly, it is not a good choice for splitting the data in any branch of the tree. In fact, using only the more discriminating attributes such as Age, Student, and Credit, the decision tree can correctly classify all 14 examples in the training set. Therefore, the additional use of Income is unnecessary and does not bring additional value to the decision making in this case.

***Predicted results:***

Age = youth, Income = medium, Student = yes, Credit = fair is Yes

* **Bayes**

Number of samples:

* + Total: 14 samples
  + Yes: 9
  + No: 5

Calculate conditional probabilities

C = Yes (9 samples):

* + P(Age = Youth | Yes) = 2 / 9
  + P(Income = Medium | Yes) = 4 / 9
  + P(Student = Yes | Yes) = 6 / 9
  + P(Credit = Fair | Yes) = 6 / 9
  + P(Yes) = 9 / 14

C = No (5 samples):

* + P(Age = Youth | No) = 3 / 5
  + P(Income = Medium | No) = 2 / 5
  + P(Student = Yes | No) = 1 / 5
  + P(Credit = Fair | No) = 2 / 5
  + P(No) = 5 / 14

Calculate Score for each class:

* + Score (Yes)



* + Score (No)



Conclusion:

Since Score(Yes) > Score(No), the sample: Age = Youth, Income = Medium, Student = Yes, Credit = Fair belongs to class Yes

**Question 2. Classification with Decision tree and Bayes.**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Object** | **Body** | **Nationality** | **Marital status** | **Group** |
| O1 | Small | Germany | Single | A |
| O2 | Big | France | Single | A |
| O3 | Big | Germany | Single | A |
| O4 | Small | Italy | Single | B |
| O5 | Big | Germany | Maried | B |
| O6 | Big | Italy | Single | B |
| O7 | Big | Italy | Maried | B |
| O8 | Small | Germany | Maried | B |

* 1. Find the classification rules of the above decision table with:
* The condition attribute set {Body, Nationality, Marital status}
* Classification attribute {Group}
  1. If the following 2 objects are belong to any class? If so, which class are they belong to?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| O9 | Big | Italy | Single | ? |
| O10 | Small | Italy | Maried | ? |

**My answer:**

**a. Find the classification rules**

***Comments from the data:***

* Group A includes only Single and non-Italian nationalities.
* Group B includes the remaining subjects, especially:
  + Italian nationality.
  + Married status.

***The classification rules are drawn:***

* Rule 1: If Marital Status = Single và Nationality ≠ Italy → Group = A (applies O1, O2, O3)
* Rule 2: If Nationality = Italy → Group = B (applies to O4, O6, O7)
* Rule 3: If Marital Status = Maried → Group = B (applies to O5, O7, O8)

**b. Classified by Naïve Bayes**

***Prior probability:***

* P(A) = 3/8
* P(B) = 5/8

***Calculate conditional probability for each attribute:***

* Group A
  + Attribute Body

P(Body=Big | A) = 2/3

P(Body=Small | A) = 1/3

* + Attribute Nationality

P(Nationality=Germany | A) = 2/3

P(Nationality=France | A) = 1/3

* + Attribute Marital Status

P(Marital=Single | A) = 3/3 = 1

* Với Group B
  + Attribute Body

P(Body=Big | B) = 3/5

P(Body=Small | B) = 2/5

* + Attribute Nationality

P(Nationality=Germany | B) = 3/5

P(Nationality=France | B) = 2/5

* + Attribute Marital Status

P(Marital=Single | ) = 2/5

P(Marital=Maried | ) = 3/5

***Calculate posterior probability:***

* O9: Big – Italy – Single
  + P(A|O9) ∝ P(A) × P(Big|A) × P(Italy|A) × P(Single|A)

= (3/8) × (2/3) × (0) × (1) = 0

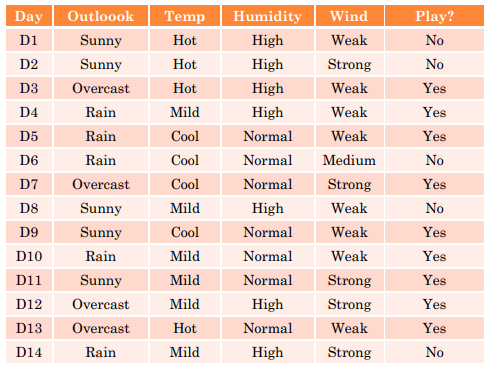
* + P(B|O9) ∝ (5/8) × (3/5) × (3/5) × (2/5) = 0.045

→ O9 belongs to Group B

* O10: Small – Italy – Maried
* P(A|O10) ∝ (3/8) × (1/3) × (0) × (0) = 0
* P(B|O10) ∝ (5/8) × (2/5) × (3/5) × (3/5) = 0.045

→ O10 belongs to Group B

**Question 3. Given data set**

****

If X= (O = Sunny, Temp = Hot, Humidity = Normal, Wind = Weak),

Which answer should be? Yes or No

**My answer:**

***Calculate the prior probability P(Play=Yes) and P(Play=No)***

Total samples = 14

* Play = Yes: 9 → P(Yes)= 9/14 ≈ 0.643
* Play = No: 5 → P(No)=5/14 ≈ 0.357

***Conditional authenticity for each specific***

For **Play = Yes** (9 samples):

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Value | Count in Yes | Probability |
| Outlook | Sunny | 2 | 2/9 ≈ 0.222 |
| Temp | Hot | 2 | 2/9 ≈ 0.222 |
| Humidity | Normal | 6 | 6/9 ≈ 0.667 |
| Wind | Weak | 6 | 6/9 ≈ 0.667 |

P(X∣Yes)=0.222⋅0.222⋅0.667⋅0.667≈0.0218

P(Yes∣X)∝P(Yes)⋅P(X∣Yes)=0.643⋅0.0218≈0.0140

For Play = No (5 samples)

|  |  |  |  |
| --- | --- | --- | --- |
| Feature | Value | Count in No | Probability |
| Outlook | Sunny | 3 | 3/5 = 0.6 |
| Temp | Hot | 2 | 2/5 = 0.4 |
| Humidity | Normal | 1 | 1/5 = 0.2 |
| Wind | Weak | 2 | 2/5 = 0.4 |

P(X∣No)=0.6⋅0.4⋅0.2⋅0.4=0.0192

P(No∣X)∝P(No)⋅P(X∣No)=0.357⋅0.0192≈0.0068

***Compare and choose labels***

P(Yes∣X)=0.0140>P(No∣X)=0.0068⇒Prediction: Yes