1. According to the table and answer the questions below

|  |  |  |  |
| --- | --- | --- | --- |
|  | Play-basketball | Not play-basketball | Total |
| Eat-cereal | 100 | 50 | 150 |
| Not eat-cereal | 150 | 300 | 450 |
| Total | 250 | 350 | 600 |

1. Calculate the conditional probability of not playing-basketball among people who eat Eat-cereal.

Ans: 50/150 = 1/3

1. Calculate the conditional probability of playing-basketball among those who do not eat-cereal.

Ans: 150/450 = 1/3

1. 下表為天氣狀況與張三、李四與王五帶傘出門狀況的統計表。請根據統計資料回答以下問題。

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| No | 天氣 | 張三帶傘 | 李四帶傘 | 王五帶傘 |
| 1 | 晴 | 否 | 否 | 是 |
| 2 | 晴 | 否 | 否 | 否 |
| 3 | 陰 | 是 | 是 | 是 |
| 4 | 陰 | 是 | 是 | 否 |
| 5 | 陰 | 是 | 是 | 否 |
| 6 | 雨 | 是 | 是 | 是 |
| 7 | 陰 | 是 | 否 | 否 |
| 8 | 雨 | 是 | 是 | 是 |
| 9 | 晴 | 否 | 否 | 否 |
| 10 | 晴 | 否 | 否 | 是 |

1. 在未觀察當天天氣的情況，且不知道張三、李四、王五帶傘情況之下，分析者該如何用利用歷史資料，來描述當天天氣的狀況？（即求出晴天、陰天、雨天的機率，請注意，機率在貝氏理論底下即為信心 Confidence）

Ans:

P(晴) = 4/10 ; P(陰) = 4/10 ; P(雨) = 2/10

1. 承上題，若分析者已知張三帶傘出門，則其對天氣描述狀況應修正為何？（三種天氣的條件機率）

Ans:

P(晴天 | 張三帶傘) = P(晴天 且 張三帶傘)/P(張三帶傘) = (0/10)/(6/10) = 0

P(陰天 | 張三帶傘) = P(陰天 且 張三帶傘)/P(張三帶傘) = (4/10)/(6/10) = 2/3

P(雨天 | 張三帶傘) = P(雨天 且 張三帶傘)/P(張三帶傘) = (2/10)/(6/10) = 1/3

1. 承題(1)，若分析者已知李四帶傘出門，則其對天氣描述狀況應修正為何？

Ans:

P(晴天 | 李四帶傘) = P(晴天 且 李四帶傘)/P(李四帶傘) = (0/10)/(5/10) = 0

P(陰天 | 李四帶傘) = P(陰天 且 李四帶傘)/P(李四帶傘) = (3/10)/(5/10) = 3/5

P(雨天 | 李四帶傘) = P(雨天 且 李四帶傘)/P(李四帶傘) = (2/10)/(5/10) = 2/5

1. 承題(1)，若分析者已知王五帶傘出門，則其對天氣描述狀況應修正為何？

Ans:

P(晴天 | 王五帶傘) = P(晴天 且 王五帶傘)/P(王五帶傘) = (2/10)/(5/10) = 2/5

P(陰天 | 王五帶傘) = P(陰天 且 王五帶傘)/P(王五帶傘) = (1/10)/(5/10) = 1/5

P(雨天 | 王五帶傘) = P(雨天 且 王五帶傘)/P(王五帶傘) = (2/10)/(5/10) = 2/5

1. 請計算利用張三、李四與王五「有帶傘」的條件來預測天氣狀況的information gain（不需要算沒帶傘），並指出誰對預測天氣最沒幫助（最沒有貢獻）。

Ans:

切之前: - (4/10)xlog2(4/10) - (4/10) xlog2 (4/10) - (2/10)x log2 (2/10) = 1.5219

張三: - 0x log20 - (2/3)x log2 (2/3) - (1/3) xlog2 (1/3) = 0.9182

李四: - 0x log2 - (3/5) xlog2 (3/5) - (2/5)x log2 (2/5) = 0.9709

王五: - (2/5)x log2 (2/5) - (1/5)x log2 (1/5) - (2/5) xlog2 (2/5) = 1.5219

* Gain三: 1.5219 – 0.9182 = 0.6037
* Gain四: 1.5219 – 0.9709 = 0.5510
* Gain五: 1.5219 – 1.5219 = 0(最少,超沒貢獻)

1. 承上題，假設已知張三帶傘，李四沒帶傘，試比較在這條件下，三種天氣的條件機率，最高者為多少？

Ans:

P(晴天 | 張三帶&李四沒帶) = P(張三帶&李四沒帶 | 晴天) x P(晴天)/P(張三帶&李四沒帶)

* 0 = 0/4 x 4/10 / P(E)

P(陰天 | 張三帶&李四沒帶) = P(張三帶&李四沒帶 | 陰天) x P(陰天)/P(張三帶&李四沒帶)

* 1 = 1/4 x 4/10 / P(E)

P(雨天 | 張三帶&李四沒帶) = P(張三帶&李四沒帶 | 雨天) x P(雨天)/P(張三帶&李四沒帶)

* 0 = 0/2 x 2/10 / P(E)

P(E) = 0 x 4/10 + 1/4 x 4/10 + 0 x 2/10 = 1/10

* 晴天跟雨天都是0
* 陰天機率100%

1. 承上題，假設已知張三帶傘，李四沒帶傘，試以Naïve-Bayes Classifier，推測是哪種天氣。(使用Laplacian correction 修正)

Ans:

P(晴天 | 張三帶傘&李四沒帶) α P(晴天) x P(張三帶傘 | 晴天) x P(李四沒帶 | 晴天)

= 4/10 x 0/4 x 4/4 = 0

* Laplacian correction => 4/10 x 1/6 x 5/6 = 0.056

P(陰天 | 張三帶傘&李四沒帶) α P(陰天) x P(張三帶傘 | 陰天) x P(李四沒帶 | 陰天)

= 4/10 x 4/4 x 1/4

* Laplacian correction => 4/10 x 5/6 x 2/6 = 0.111

P(雨天 | 張三帶傘&李四沒帶) α P(雨天) \* P(張三帶傘 | 雨天) \* P(李四沒帶 | 雨天)

= 2/10 x 2/2 x 0/2

* Laplacian correction => 2/10 x 3/4 x 1/4 = 0.0375

🡺 推測為陰天~

-----------------------------------------------------------------------------------------------------------------

classification\_2程式:

ROC 曲線，並計算 AUC 值(曲線下的面積)

* ROC曲線和AUC常用於評估二分模型的效能。
* ROC曲線展示了在不同閾值下模型的表現，而AUC提供了一個單一的數值指標，用於整體評估模型效能。
* 通常情況下，ROC曲線越靠近左上角，AUC值越高，表示模型效能越好。
* ROC和AUC都不受類別不平衡的影響，因此適用於評估不同類別分佈下的模型效能。











