1. 某房仲業者蒐集1,000位35歲的工程師的婚姻狀態與購置房屋不動產的統計資料，其中550位工程師已婚，且共有250位已購買不動產，另外450位單身的工程師中，有300位尚未購買不動產，所有檢驗統計資料列於下表。

購買房屋不動產與婚姻狀態之統計表

|  |  |  |  |
| --- | --- | --- | --- |
| 婚姻狀態  購買不動產 | 單身 | 已婚 | 合計 |
| 已購買（） | 150 | 250 | 400 |
| 未購買（） | 300 | 300 | 600 |
| 合計 | 450 | 550 | 1000 |

1. 請計算來自已婚的工程師中已購買房屋不動產的條件機率。
2. 請計算來自單身的工程師中尚未購買房屋不動產的條件機率。
3. 若今天有一位已婚的工程師，請預測該位工程師是否已經購買房屋不動產？

P(單身) = 450/1000 = 0.45

P(已婚) = 550/1000 = 0.55

P(已購) = 400/1000 = 0.4

P(未購) = 600/1000 = 0.6



P(單身∩已購) = 0.45 \* 0.4 = 0.18

P(單身∩未購) = 0.45 \* 0.6 = 0.27

P(已婚∩已購) = 0.55 \* 0.4 = 0.22

P(已婚∩未購) = 0.55 \* 0.6 = 0.33

P(單身∣已購) = 150/400 = 0.375

P(已婚∣已購) = 250/400 = 0.625

P(單身∣未購) = 300/600 = 0.5

P(已婚∣未購) = 300/600 = 0.5

P(已購∣單身) = 150/450 = 0.333333

P(已購∣已婚) = 250/550 = 0.454545

P(未購∣單身) = 300/450 = 0.666666

P(未購∣已婚) = 300/550 = 0.545454

(1) 62.5%

(2) 50%

(3) 否

P(已婚∣已購) \* P(已購) = 0.625 \*0.4 = 0.25

P(已婚∣未購) \* P(未購) = 0.5 \* 0.6 = 0.3

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 | 王五帶傘 | 李四帶傘 | 張三帶傘 |
| 1 | 晴 | 是 | 否 | 否 |
| 2 | 晴 | 否 | 否 | 否 |
| 3 | 陰 | 是 | 否 | 是 |
| 4 | 陰 | 否 | 是 | 是 |
| 5 | 雨 | 否 | 是 | 是 |
| 6 | 雨 | 是 | 是 | 是 |
| 7 | 陰 | 否 | 否 | 是 |
| 8 | 雨 | 是 | 是 | 是 |
| 9 | 晴 | 否 | 否 | 否 |
| 10 | 晴 | 是 | 否 | 否 |

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）

P(晴) = 4/10 = 0.4 = 40%

P(陰) = 3/10 = 0.3 = 30%

P(雨) = 3/10 = 0.3 = 30%

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）

P(天氣 = 晴∣張三 = 是) = 0/6 = 0%

P(天氣 = 陰∣張三 = 是) = 3/6 = 0.5 = 50%

P(天氣 = 雨∣張三 = 是) = 3/6 = 0.5 = 50%

P(晴∣張三) = P(晴) \* P(天氣 = 晴∣張三 = 是) = 0.4 \* 0 = 0

P(陰∣張三) = P(陰) \* P(天氣 = 陰∣張三 = 是) = 0.3 \* 0.5 = 0.15

P(雨∣張三) = P(雨) \* P(天氣 = 雨∣張三 = 是) = 0.3 \* 0.5 = 0.15

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？

P(天氣 = 晴∣李四 = 是) = 0/4 = 0%

P(天氣 = 陰∣李四 = 是) = 1/4 = 0.25 = 25%

P(天氣 = 雨∣李四 = 是) = 3/4 = 0.75 = 75%

P(晴∣李四 + 張三 ) = 0.4 \* 0 \* 0 = 0

P(陰∣李四 + 張三 ) = 0.3 \*0.25 \* 0.5 = 0.0375

P(雨∣李四 + 張三 ) = 0.3 \* 0.75 \* 0.5 = 0.1125

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？

P(天氣 = 晴∣王五 = 是) = 2/5 = 0.4 = 40%

P(天氣 = 陰∣王五 = 是) = 1/5 = 0.2 = 20%

P(天氣 = 雨∣王五 = 是) = 2/5 = 0.4 = 40%

P(晴∣李四 + 張三 + 王五) = 0.4 \* 0 \* 0 \* 0.4 = 0

P(陰∣李四 + 張三 + 王五) = 0.3 \*0.25 \* 0.5 \* 0.2 = 0.0075

P(雨∣李四 + 張三 + 王五) = 0.3 \* 0.75 \* 0.5 \* 0.2 = 0.0225

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），亂度請利用自然對數 e 為底。

王五: I(5,5) = 

= 

李四: I(4,6) = )

= 

張三: I(6,4) = 

= 

Ans: 王五 對天氣預測最沒幫助

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？

P(晴) = 4/10 = 0.4 = 40%

P(陰) = 3/10 = 0.3 = 30%

P(雨) = 3/10 = 0.3 = 30%

P(天氣 = 晴∣張三 = 是) = 0/6 = 0%

P(天氣 = 陰∣張三 = 是) = 3/6 = 0.5 = 50%

P(天氣 = 雨∣張三 = 是) = 3/6 = 0.5 = 50%

P(天氣 = 晴∣李四 = 否) = 4/6 = 0.6666 = 66.66%

P(天氣 = 陰∣李四 = 否) = 2/6 = 0.3333 = 33.33%

P(天氣 = 雨∣李四 = 否) = 0/6 = 0%

P(晴∣張三 + 李四 ) = 0.4 \* 0 \* 0.6666 = 0

P(陰∣張三 + 李四 ) = 0.3 \* 0.5 \* 0.3333 = 0.049995

P(雨∣張三 + 李四 ) = 0.3 \* 0.5 \* 0 = 0

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: 陰天

P(晴) = 4/10 = 0.4 = 40%

P(陰) = 3/10 = 0.3 = 30%

P(雨) = 3/10 = 0.3 = 30%

P(天氣 = 晴∣張三 = 是) = 1/9 = 0.11111 = 11.11%

P(天氣 = 陰∣張三 = 是) = 4/9 = 0.4444 = 44.44%

P(天氣 = 雨∣張三 = 是) = 4/9 = 0.4444 = 44.44%

P(天氣 = 晴∣李四 = 否) = 5/9 = 0.5555 = 55.55%

P(天氣 = 陰∣李四 = 否) = 3/9 = 0.3333 = 33.33%

P(天氣 = 雨∣李四 = 否) = 1/9 = 0.1111 = 11.11%

P(晴∣張三 + 李四 ) = 0.4 \* 0.1111 \* 0.5555 = 0.02468642

P(陰∣張三 + 李四 ) = 0.3 \* 0.4444 \* 0.3333 = 0.04443556

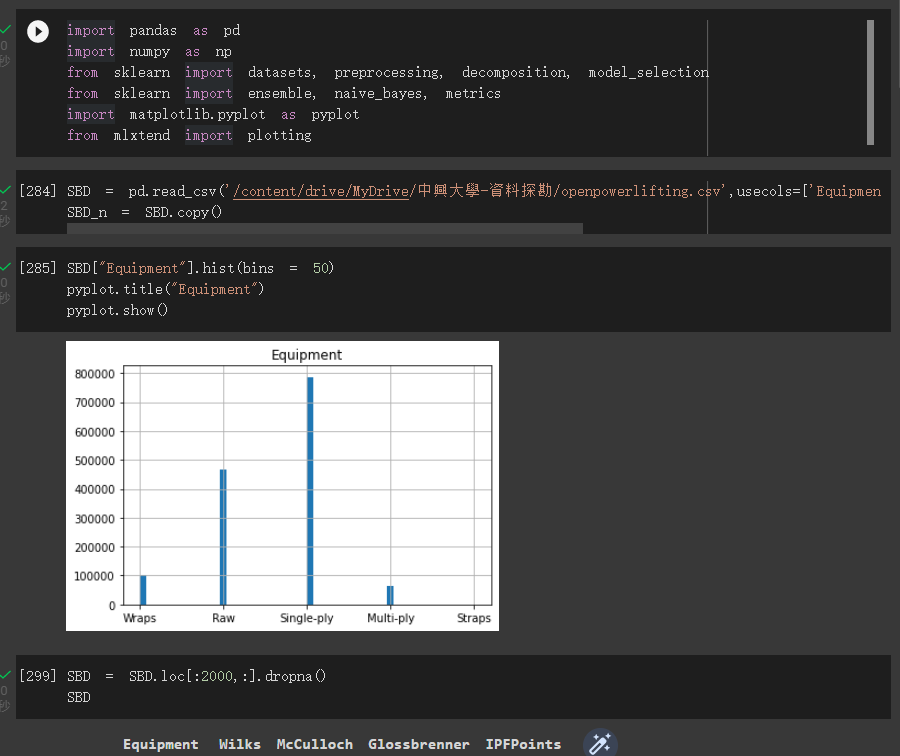
P(雨∣張三 + 李四 ) = 0.3 \* 0.4444 \* 0.1111 = 0.01481185

註：

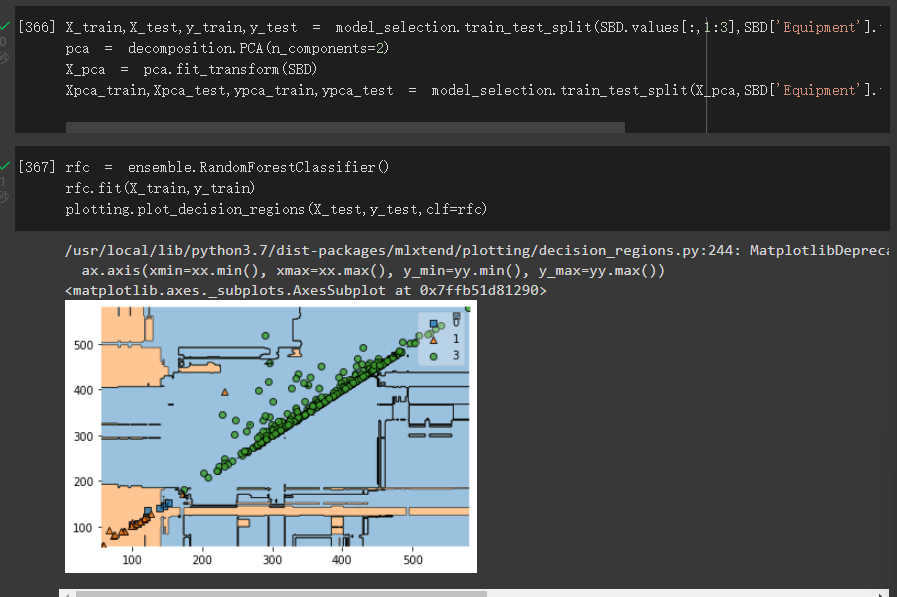
第 (5) 題未切割前亂度：-4/10\*log(4/10) - 3/10\*log(3/10) - 3/10\*log(3/10) = 1.088 （可利用下方python程式碼算出）

from math import log

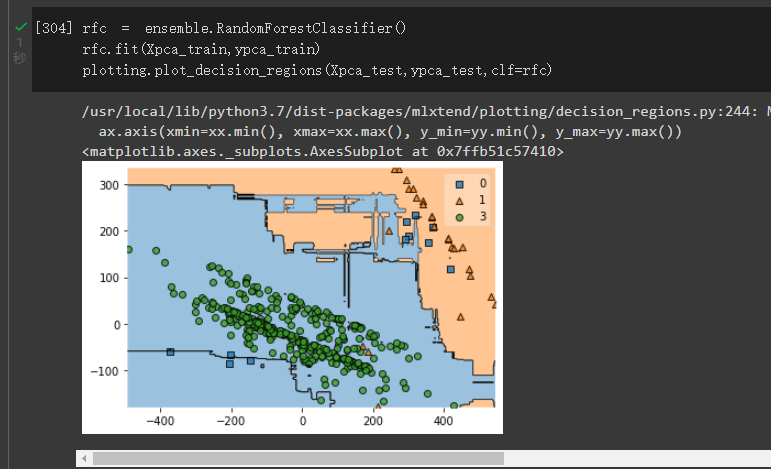
-4/10\*log(4/10) - 3/10\*log(3/10) - 3/10\*log(3/10)



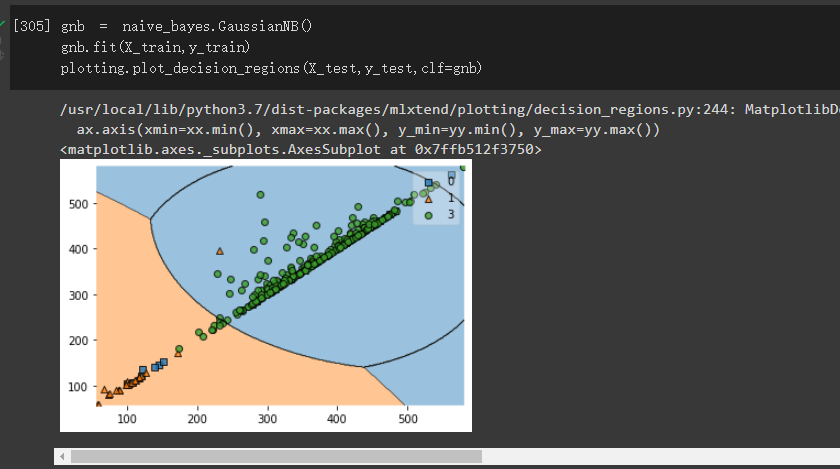




**PCA**







**PCA**

