Data Science HW5

Prof. M.-S. Chen & Prof. M.-L. Lo

TA Office hours

- Monday 14:00-15:00 @ 博理603
- Contact
  - email: ntuds.ta@gmail.com

繳交方式

- Deadline: 2022/12/9 Fri. 23:59
- 上傳格式 : PDF, to COOL

HW5 扣分規則

- 遲交
  - 此次作業 0 分 
- 格式錯誤
  - 此次作業 -10 分

Problem

Customer ID	Gender	Car Type	Shirt Size	Class
1	M	Family	Small	C0
2	M	Family	Medium	C0
3	M	Family	Extra Large	C0
4	M	Sports	Extra Large	C0
5	M	Family	Large	C0
6	M	Family	Extra Large	C0
7	M	Luxury	Extra Large	C0
8	F	Family	Small	C0
9	F	Family	Medium	C0
10	F	Luxury	Large	C0
11	F	Luxury	Large	C0
12	M	Sports	Medium	C1
13	M	Sports	Large	C1
14	M	Sports	Medium	C1
15	F	Sports	Small	C1
16	F	Luxury	Small	C1
17	F	Luxury	Small	C1
18	F	Sports	Medium	C1
19	F	Luxury	Medium	C1
20	F	Luxury	Medium	C1

Problem

1 (40%) For the data in the above table, please use gini index (two way split) to derive and draw the resulting decision tree.
(Please also show every gini index of candidates and the rule you choose at every step.)

2 (30%) For the data in the above table, use Naïve Bayes classifier to classify an input tuple: (Gender=M, Car Type=Sports, Shirt Size=Medium)
(Please show your procedure)

3.(30%) Drive the hyperplane by the SVM procedure taught in the class.
(Please show your procedure)

Positive samples, y = 1: (4, 3), (4, 8), (7, 2)
Negative samples, y = -1: (-1, -2), (-1, 3), (2, -1), (2, 1).

溫馨提醒

常見問題

❖ 第一題 (也可以用程式計算)

1. 如果遇到gini-index數值相同，就取隨便一個都可以。
2. 每一個node請針對單一個feature使用Two-way split的分法 (請參考第10週上課講義p.45)
3. 請畫出最後的樹，並標注每個node的gini-index, leaf node分出的類別 (如果最後的leaf node已無法再split, 也請註明清楚)

❖ 第二題

1. 請標明計算過程

❖ 第三題 (也可以用程式計算)

在report中一定要註明

1. objectives & constraints
2. 寫出兩個support vectors
3. 列出計算過程
4. 寫出 w, b, and hyperplane ($y=w^Tx+b$)