

# Data Science

---

**Prof M.-S. Chen**

**Prof M.-L Lo**

# TA Office ours

- ❖ Every Friday AM11:00-12:00 博理603
- ❖ Contact
  - email: [datascience2020fall@gmail.com](mailto:datascience2020fall@gmail.com)
  - FB QA [discuss community](#)



# DS HW5 - Classification

# 繳交方式

- ◆ Deadline: **2020/12/8 Tue. 23:59**
- ◆ 上傳格式 : **PDF**, to ceiba

# HW5 扣分規則

- ❖ 遲交
  - 此次作業 **0** 分 :<
- ❖ 格式錯誤
  - 此次作業 **-20** 分

**Problem**

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

想複製表格的點這邊: [\[here\]](#)

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 **show steps**, every gini index of candidates and which rule you choose; also draw out the tree).

2 (30%) For the data in the above table, use Naïve Bayes classifier to classify an input tuple.

with (Gender=F, Car Type=Luxury, Shirt Size=Large).

(Please **show your procedure**).

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

(Please **show your procedure**).

Positive examples,  $y = 1$ : (4, 3), (7, 2), (4, 8).

Negative examples,  $y = -1$ : (2, 1), (2, -1), (-1, 3), (-1, -2).



**溫馨提醒**

# decicion tree

- ◆ 如果可以的話請各位第一小提盡量寫程式去計算, 用手算會算到發瘋喔!