

Data Science 2019
HW5 - Classification
Due: 2019/12/04 09:00 a.m.

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	Family	Large	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	Family	Large	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

1. 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) [40%]
2. 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) [30%]
3. Drive the hyperplane $y = wx + b$ by the SVM procedure taught in the class. [30%]
Positive examples, $y = 1$: (4, 3), (7, 2), (4,8)
Negative examples, $y = -1$: (2, 1), (2, -1), (-1, 3), (-1, -2)
You may choose either one to answer:
 1. It's not necessary to calculate the exact w and b , but you need to find out the boundary of w (with your procedure), and explain how to find w conceptively.
 2. Find two support vectors from graph (should be obvious), and find out hyperplane $y = wx + b$.

Submission

- Write down the process you get the answer as clear as possible.
- Please submit your homework in **PDF** format. (no naming rule, no zip ... etc.)
- You may submit handwriting by scanning or capturing, be sure that it's clear enough.
- Wrong submission format will get 10 points penalty.
- No plagiarism.
- Submission deadline: **2019/12/05 09:00 a.m.**
- Accept late submission for 2 days after deadline.
- Late submission penalty is 20% per day (start from 9:00a.m.)