Problem Set 2

*Start Date: Oct. 19, 2021
*Due Date: Nov. 02, 2021

Notes:

- Each homework problem must be done individually.
- Please make sure your full name, your student ID, and the assignment number appears at the top of each file you submit.
- Please answer questions in the order they are assigned.
- Please submit a **PDF** version. Homework may be typed or handwritten but handwritten submissions need to be legible.
- Please submit on the Web Learning platform. Do NOT Email or Wechat your assignment to the instructor or TAs.
- No late homework is accepted. No exceptions.
- You can answer the question using English or Chinese.

1. Given the following data, we have two features A and B, three labels $C1$, $C2$ and $C3$. Try to calculate the Information Gain(IG) of A and B, give the decision tree basing on the IG your get.

Label	A	B
C3	1	2
C3	1	1
C1	1	2
C1	1	2
C1	2	2
C2	2	1
C2	3	1
C2	3	1
C2	3	1
C2	3	2

2. Write down the primal form of linear SVM for cases when the samples are not linearly separable, write down the Lagrangian of the problem, and derive the dual problem. Derive $|w_0|^2$ at the solution, and show that $a_i = c$ for samples that are not correctly classified. (Hint: you can refer to website)