School of Computing and Information Systems
The University of Melbourne
COMP90049 Introduction to Machine Learning (Semester 1, 2023)
Week 7

1. Consider a Naive Bayes model trained using the following familiar weather dataset:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
A	s	h	n	F	N
B	s	h	h	T	N
C	o	h	h	F	Y
D	r	m	h	F	Y
E	r	c	n	F	Y
F	r	c	n	T	N

Suppose that you made additional observations of days and their features. But you don't have the label for the PLAY in these days:

ID	*Outl*	*Temp*	*Humi*	*Wind*	PLAY
G	o	m	n	T	?
H	s	m	h	F	?

How could you incorporate this information into your Naïve Bayes model without manually annotating the labels? If necessary, recompute your model parameters.

2. What is the main assumption of self-training? What is the main assumption of Active Learning?

3. (a) Describe the rationale and key principles behind the Query-by-Committee algorithm. (b) Use QBC to determine the instance that our active learner would select first in the following scenario.

classifier	Instance 1			Instance 2			Instance 3		
	y_1	y_2	y_3	y_1	y_2	y_3	y_1	y_2	y_3
C_1	0.2	0.7	0.1	0.2	0.7	0.1	0.6	0.1	0.3
C_2	0.1	0.3	0.6	0.2	0.6	0.2	0.21	0.21	0.58
C_3	0.8	0.1	0.1	0.05	0.9	0.05	0.75	0.01	0.24
C_4	0.3	0.5	0.2	0.1	0.8	0.1	0.1	0.28	0.62