School of Computing and Information Systems
The University of Melbourne
COMP90049 Introduction to Machine Learning (Semester 2, 2022)
Week 2: Workshop

Considering the following problems:

(i) Skin cancer screening test

(ii) Building a system that guesses what the weather (temperature, precipitation, etc.) will be like tomorrow

(iii) Predicting products that a customer would be interested in buying, based on other purchases that customer has previously made

1. Identify the "concept" we might attempt to "learn" for each problem (Task Identification)

2. For each problem-task, identify what the "instances" and "attributes" might consist of (Choosing the Data Representative)

3. For each problem-task, conjecture whether a typical strategy is likely to use "supervised" or "unsupervised" Machine Learning (Picking a Suitable Model)

4. [OPTIONAL] For each problem-task, consider how easy or difficult it would be to make a model that generalizes to new cases. For example, could you predict the weather in any city in the world, or just in one specific city?

5. [OPTIONAL] What kinds of assumptions might a machine learning model make when tackling these problems?

6. What is **discretisation**, and where might it be used? Discretise attribute C of the following dataset according to the given methods (breaking ties where necessary).

   (i) Equal width

   (ii) Equal frequency

   (iii) k-means

ID	A (°C)	B (mm)	C (hPa)	CLASS
1	22.5	4.6	1021.2	AUT
2	16.7	21.6	1027.0	AUT
3	29.6	0.0	1012.5	SUM
4	33.0	0.0	1010.4	SUM
5	13.2	16.4	1019.5	SPR
6	14.9	8.6	1016.4	SPR
7	18.3	7.8	995.4	WIN
8	16.0	5.6	1012.8	WIN