

COMP4432 Machine Learning

Tutorial Questions on Decision Tree

1. Given the following table.

<i>Parcel ID</i>	<i>Origin</i>	<i>Destination</i>	<i>Type</i>	<i>Weight</i>
1	HK	HK	Parcel	Light
2	Kln	Kln	Letter	Light
3	NT	Kln	Letter	Light
4	HK	HK	Parcel	Heavy
5	Kln	Kln	Parcel	Light
6	NT	NT	Letter	Light
7	HK	HK	Letter	Light
8	Kln	Kln	Parcel	Heavy
9	Kln	Kln	Letter	Light
10	HK	HK	Letter	Light
11	HK	HK	Parcel	Heavy
12	Kln	Kln	Letter	Light
13	HK	HK	Letter	Light
14	Kln	Kln	Parcel	Light
15	HK	NT	Parcel	Heavy
16	NT	Kln	Letter	Light
17	HK	NT	Letter	Light
18	Kln	HK	Parcel	Light
19	HK	NT	Parcel	Heavy
20	HK	HK	Parcel	Light
21	Kln	Kln	Letter	Light
22	Kln	HK	Parcel	Heavy
23	Kln	Kln	Letter	Light
24	Kln	Kln	Letter	Light
25	HK	HK	Parcel	Light

Construct a decision tree, based on information gain, to classify the type of courier services (cf. column *Type*). You may assume that the first 20 records are available for model construction and the remaining 5 records are used to validate your answer.

2. You are given the following 40 2-D points belonging to two classes, i.e., blue class and orange class. They are integer-aligned, i.e., (0,0; blue), (2,0; blue), (4,0; blue), (6,0; blue), (8,0; blue), (10,0; orange), (12,0; orange), (14,0; orange), (0,2; blue), ... , (14,8; orange).

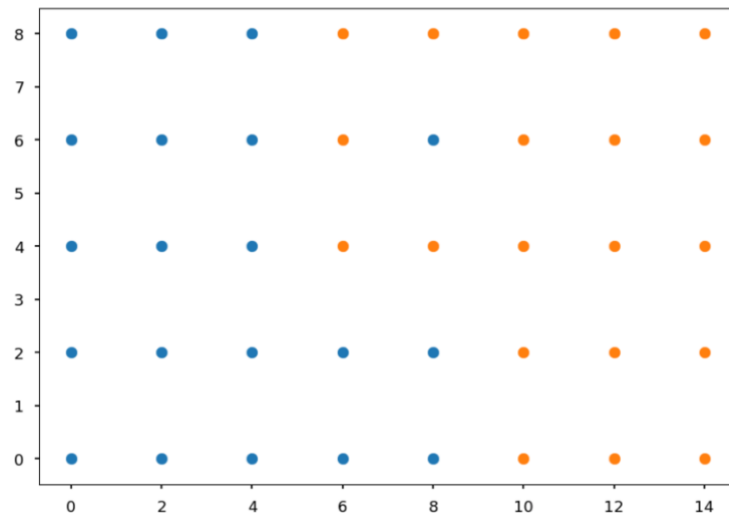


Fig.2 2D Dataset for Q.2

- According to the idea of ID3 (and referring to slide 31 of lecture notes-DT), use the available 40 points to build a decision tree with classification accuracy (training) $\geq 95\%$. Comprehensive steps of your computation are NOT required but indicative ones are expected.
- Draw the decision surface of the decision tree you build in part (a).
- If the classification accuracy (training) is now lifted to 100%, build a corresponding decision tree and draw its decision surface. Again, a concise answer is not needed.
- As a second thought, can we obtain a decision tree with 2 levels only (root \rightarrow level_1 \rightarrow leaves) and axis-parallel decision nodes for the criterion of classification accuracy (training) $\geq 95\%$?