Introduction to Artificial Intelligence
Supervised ML

Exercise 1. The table below provides a training data set containing six observations, three predictors, and one qualitative response variable.

Obs.	X_1	X_2	X_3	Y
1	0	3	0	Red
2	2	0	0	Red
3	0	1	3	Red
4	0	1	2	Green
5	-1	0	1	Green
6	1	1	1	Red

Suppose we wish to use this data set to make a prediction for Y when $X_1 = X_2 = X_3 = 0$ using K-nearest neighbors.

(a) Compute the Euclidean distance between each observation and the test point, $X_1 = X_2 = X_3 = 0$.

(b) What is our prediction with $K = 1$? Why?

(c) What is our prediction with $K = 3$? Why?

(d) If the Bayes decision boundary in this problem is highly non- linear, then would we expect the best value for K to be large or small? Why?

2. Naïf Bayes Classifier

A French Region wants to understand the investment differences in several towns and villages. To study this problem, several criteria was taking into account:

- Existence of a Train Station: yes/no
- Status: Town/Village
- Distance to Paris: close/far
- Investment level: small/medium/high

Station	**Status**	**Distance**	**Investment**
No	Village	Close	Medium
Yes	Town	Close	High
No	Village	Far	Small
Yes	Village	Close	High
No	Town	Far	Small
Yes	Town	Far	Medium
No	Town	Close	High

2.1 Use the Bayes Classifier to compute build the model (compute all the probabilities) for each class and for each variable.

2.2 If a Village has a train station and it is situated far from Paris, predict the investment level.

3. Decision Tree

Use the same previous dataset and problem by applying the Decision Tree.

3.1 Build a Decision Tree using like Class the attribute Investment computing the Cross-entropy.

3.2 If a Village has a train station and it is situated far from Paris, predict the investment level.