The University of Adelaide
School of Computer Science

Artificial Intelligence
Tutorial 3

Question 1
Figure 1 shows a scatter plot of points $p_1, p_2, \ldots, p_{10}$ with their class labels. The plot also includes the testing point z.

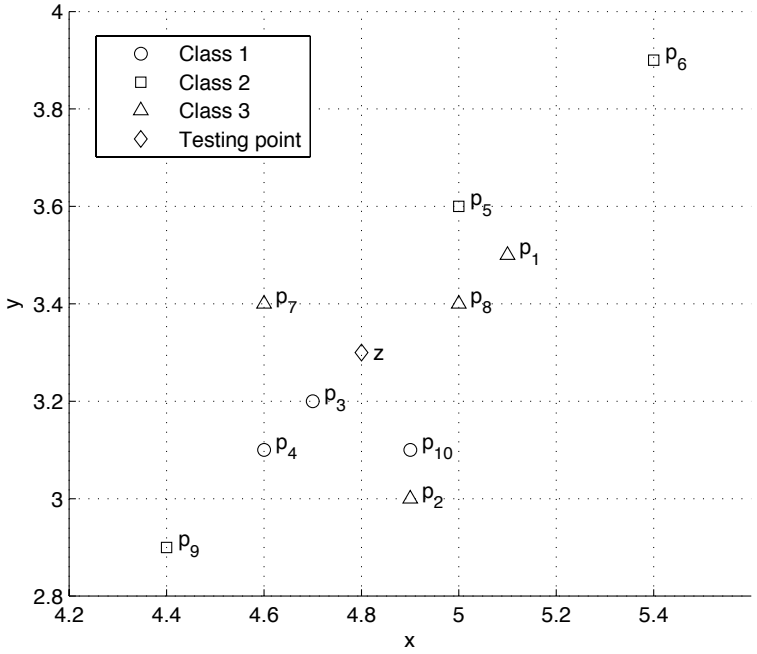

Figure 1: Scatter plot of points with class labels.

The precise coordinates of all the points are as follows:

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	z
x	5.1	4.9	4.7	4.6	5.0	5.4	4.6	5.0	4.4	4.9	4.8
y	3.5	3.0	3.2	3.1	3.6	3.9	3.4	3.4	2.9	3.1	3.3

Classify the testing point z using K nearest neighbours with $K = 1, 3, 4, 5$ and 7.

Question 2
Using the training data $X = \{p_1, p_2, \ldots, p_{10}\}$ in Figure 1,

1. Construct a Kd-tree with a bucket-size of 1, i.e., each leaf node must contain at least 1 point. Indicate clearly the branching criterion at each node.

2. In Figure 1, show clearly the spatial partitioning corresponding to the constructed Kd-tree.

3. Search the Kd-tree for the nearest neighbour of testing point z, indicating clearly on the Kd-tree the nodes that are visited.

Question 3
Given a set of points in 2D and a rectangle of known size and position, we wish to search for the points that lie in the rectangle; see Figure 2.

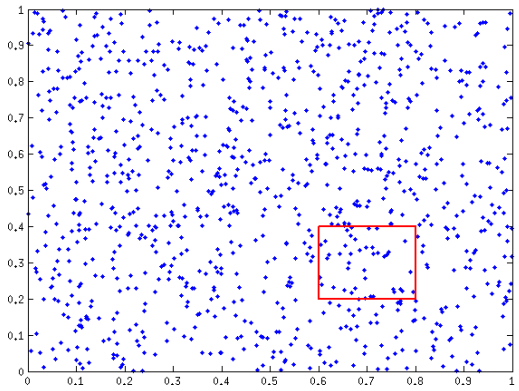

Figure 2: Find the points that lie in the rectangle.

Explain how you would modify the Kd-tree approach to enable this type of search to be accomplished quickly. Specifically, describe

1. What extra information that you need to store in the Kd-tree data structure.

2. How the Kd-tree search algorithm should be modified to enable this task.

Question 4 (Question 18.15 in AIMA 3ed)
Suppose a 7-nearest-neighbors regression search returns $\{7,6,8,4,5,11,100\}$ as the 7 nearest y values for a given x value. What is the value of $\hat{y}$ (the predicted output value of x) that minimizes the L_1 loss function (sum of absolute errors) on this data? There is a common name in statistics for this value as a function of the y values; what is it? Answer the same two questions for the L_2 loss function (sum of squared errors).

Question 5 (Question 20.21 in AIMA 2ed)
Consider the problem of separating N data points into positive and negative examples using a linear separator. Clearly, this can always be done for $N = 2$ points on a line of dimension $d = 1$, regardless of how the points are labelled or where they are located (unless the points are in the same place).

1. Show that it can always be done for $N = 3$ points on a plane of dimension $d = 2$, unless they are collinear.

2. Show that it cannot always be done for $N = 4$ points on a plane of dimension $d = 2$.

3. Show that it can always be done for $N = 4$ points in a space of dimension $d = 3$, unless they are coplanar.

4. Show that it cannot always be done for $N = 5$ points in a space of dimension $d = 3$.

Question 6

A compilation of the playing conditions and outcomes of matches between tennis players Federera and Nadale is given in Table 1, where the time of the match is either Morning (M), Afternoon (A) or Night (N); the type of match is either Grand Slam (G), Master (M) or Friendly; the type of court is either Grass (G), Hard (H), Clay (C) or Mixed (M); and the outcome is a Federera (F) win or a Nadale (N) win.

Time	Match	Surface	Outcome
M	M	G	F
A	G	C	F
N	F	H	F
A	F	M	N
A	M	C	N
A	G	G	F
A	G	H	F
A	G	H	F
M	M	G	F
A	G	C	N
N	F	H	F
N	M	M	N
A	M	C	N
A	M	G	F
A	G	H	F
A	G	C	F

Table 1: A record of previous tennis matches between Federera and Nadale.

Build a decision tree that can predict the outcome of a new match, given information about the time, type of match, and surface. Show all your working to demonstrate that you are correctly using information gain as the splitting criterion.

Question 7

Figure 3 shows data pertaining to two species of Irises (Setosa and Virginica). The

data contains measurements of two attributes, sepal length and width. Our goal is to construct a decision tree based on this set of training data to predict the species of a given instance of an Iris flower.

a. What is the best value for attribute sepal length to split the data (consider only values on the grid, i.e. 4, 4.25, 4.5, ...)? What is the information gain associated with this value?

b. Following the first split using attribute sepal length, we now choose to split using attribute sepal width. What are the best values for attribute sepal width to split the *remaining* data (again, consider only values on the grid, i.e. 2.25, 2.5, 2.75, ...)? What is the information gain of each split?

c. If we alternatingly choose between the two attributes to split, how many splits are required in total so that the data is cleanly separated (i.e. all leaf nodes are pure)? Mark in Figure 3 the regions corresponding to the leaf nodes and their labels.

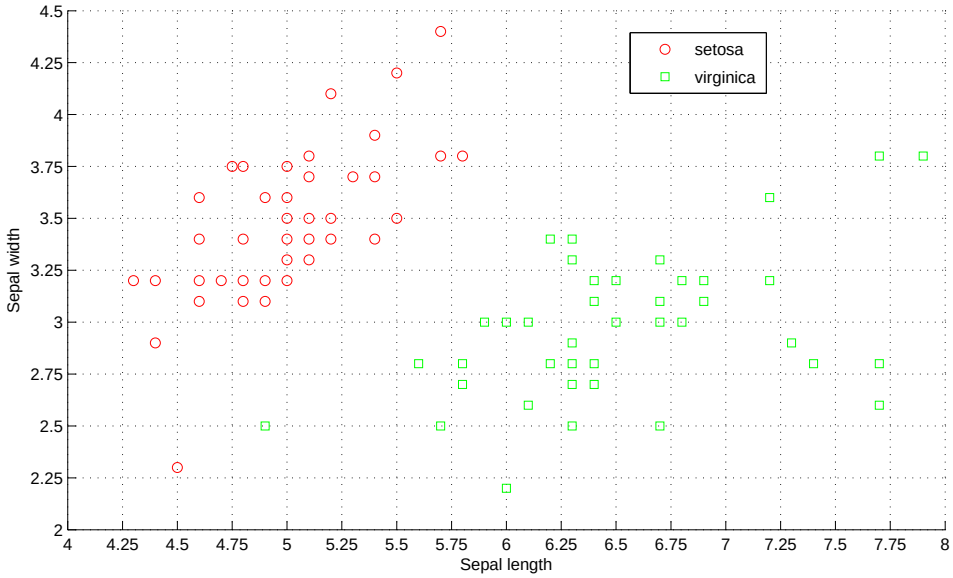

Figure 3: Iris data with two attributes.

Question 8 (Question 18.8 in AIMA 2ed)
In the recursive construction of decision trees, it sometimes happens that a mixed set of positive and negative examples remains at a leaf node, even after all the attributes have been used. Suppose that we have p positive examples and n negative examples.

- Show that the solution used by the given decision tree learning algorithm, which picks the majority classification for mixed nodes, minimizes the sum of absolute errors over the set of examples at the leaf.
- Show that using the class probability $p/(p+n)$ minimizes the sum of squared errors.

Question 9 (Question 18.10 in AIMA 2ed)
Suppose that an attribute splits the set of examples E into subsets E_k and that each subset has p_k positive examples and n_k negative examples. Show that the attribute has strictly positive information gain unless the ratio $p_k/(p_k+n_k)$ is the same for all k.

Question 10 Clustering
Figure 4 shows the a scatter plot of points $p_1, p_2, \ldots, p_{10}$. We wish to cluster the 10 points using the K-means algorithm. The two initial cluster centres are given as m_1 and m_2 as shown in the figure. The precise coordinates of all the points are as follows:

	p_1	p_2	p_3	p_4	p_5	p_6	p_7	p_8	p_9	p_{10}	m_1	m_2
x	-3.0	-1.4	-0.5	-3.5	-3.1	1.4	1.8	1.2	0.6	1.8	0.5	0.0
y	-1.9	-0.9	-1.9	-2.7	0.4	2.7	2.9	0.6	2.5	1.8	1.0	2.0

1. Calculate the revised positions of m_1 and m_2 after 1 iteration of K-means.
2. Predict the positions of m_1 and m_2 when K-means converges.

Question 11 Mean-shift
Referring to Figure 4, calculate the revised positions for

1. point p_1
2. point p_{10}

after 1 iteration of the mean-shift procedure using the Gaussian kernel with bandwidth $\sigma = 1.0$. Where will mean-shift cause p_1 and p_{10} to eventually converge to?

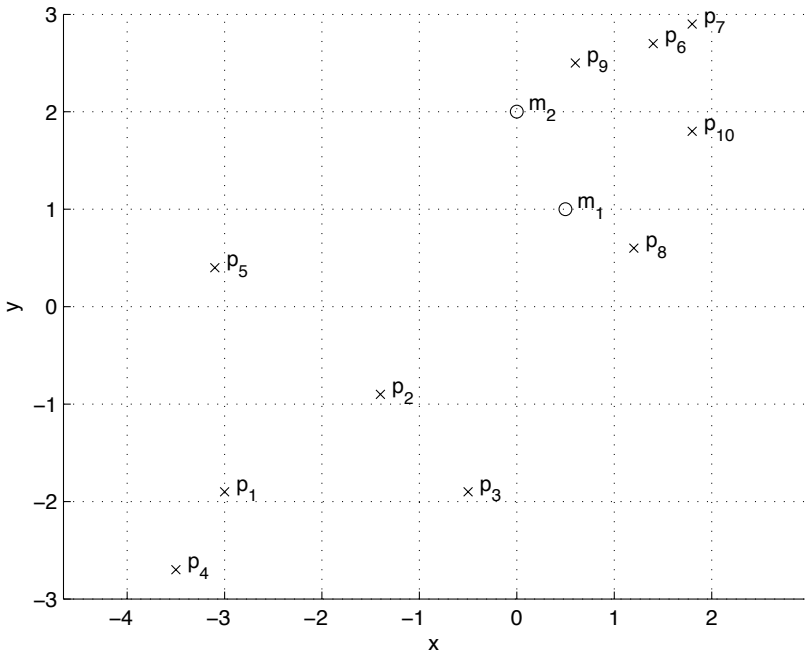

Figure 4: Scatter plot of points for clustering.