CS 540-2: Introduction to Artificial Intelligence

Homework Assignment #3 Solutions

Problem 1. [20] Unsupervised Learning by Clustering

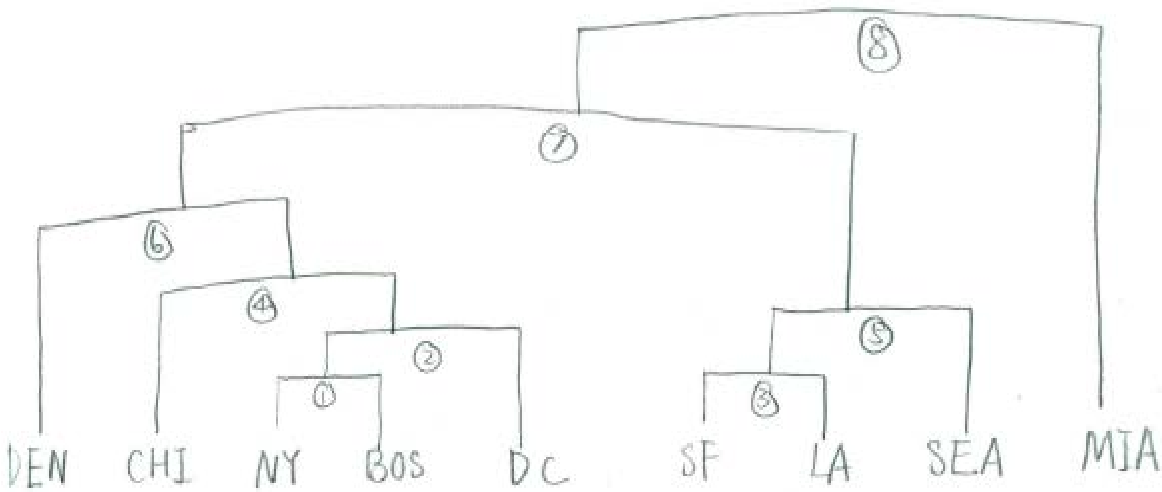
Consider the following information about *distances* in miles between pairs of 9 U.S. cities:

	BOS	NY	DC	MIA	CHI	SEA	SF	LA	DEN
BOS	0	206	429	1504	963	2976	3095	2979	1949
NY	206	0	233	1308	802	2815	2934	2786	1771
DC	429	233	0	1075	671	2684	2799	2631	1616
MIA	1504	1308	1075	0	1329	3273	3053	2687	2037
CHI	963	802	671	1329	0	2013	2142	2054	996
SEA	2976	2815	2684	3273	2013	0	808	1131	1307
SF	3095	2934	2799	3053	2142	808	0	379	1235
LA	2979	2786	2631	2687	2054	1131	379	0	1059
DEN	1949	1771	1616	2037	996	1307	1235	1059	0

The (latitude, longitude) *locations* of these cities are: BOS (42.4, 71.1), NY (41.7, 74.0), DC (38.9, 77.0), MIA (25.8, 80.2), CHI (41.9, 87.7), SEA (47.6, 122.3), SF (37.8, 122.4), LA (34.1, 118.2), and DEN (39.7, 105.0).

(a) [10] Perform (manually) **hierarchical agglomerative clustering** using *single-linkage* and the above data.

i. [8] Show the resulting dendrogram.



ii. [2] What clusters of cities are created if you want 3 clusters?

Cluster 1: DEN, CHI, NY, BOS, DC
Cluster 2: LA, SF, SEA
Cluster 3: MIA

(b) [10] Show the results of one iteration of ***k*-means clustering** assuming $k = 2$ and the initial cluster centers are defined as $c_1 = (38.0, 103.0)$ and $c_2 = (30.0, 78.0)$

i. [3] Give the list of cities in the initial 2 clusters.

Cluster 1: SEA, SF, LA, DEN
Cluster 2: BOS, NY, DC, MIA, CHI

ii. [4] Give the coordinates of the new cluster centers.

Center of Cluster 1: (39.8, 116.98)
Center of Cluster 2: (38.14,78)

iii. [3] Give the list of cities in the 2 clusters based on the new cluster centers computed in (ii).

Cluster 1: SEA, SF, LA, DEN
Cluster 2: BOS, NY, DC, MIA, CHI

Problem 2: Decision Trees [25 points]

The following table summarizes a training set containing 100 examples where each example has 3 binary attributes, *A*, *B* and *C*, and there are two class labels, $Y \in \{+, -\}$.

A	B	C	Y	
			+	−
T	T	T	5	0
F	T	T	0	20
T	F	T	20	0
F	F	T	0	5
T	T	F	0	0
F	T	F	25	0
T	F	F	0	0
F	F	F	0	25

(a) What is the **entropy** of *Y*, *H*(*Y*), as computed from these 100 examples?

$$H(Y) = -P(Y=+)\log_2 P(Y=+) - P(Y=-)\log_2 P(Y=-)$$
$$= -50/100 \log 50/100 - 50/100 \log 50/100 = 1$$

(b) What is the **conditional entropy** of *Y* given *A*? That is, compute *H*(*Y* | *A*).

$$H(Y|A) = P(A=T)H(Y|A=T) + P(A=F)H(Y|A=F)$$
$$= 25/100[-P(Y=+|A=T)\log P(Y=+|A=T) - P(Y=-|A=T)\log P(Y=-|A=T)]$$
$$+ 75/100[-P(Y=+|A=F)\log P(Y=+|A=F) - P(Y=-|A=F)\log P(Y=-|A=F)]$$
$$= 0.25[-25/25 \log 25/25 - 0/25 \log 0/25]$$
$$+ 0.75[-25/75 \log 25/75 - 50/75 \log 50/75]$$
$$= 0.25[0] + 0.75[-1/3 \log 1/3 - 2/3 \log 2/3]$$
$$= .75[-(.33)(-1.59) - (.67)(-.58)]$$
$$= 0.68$$

(c) What is the **information gain** between attribute *A* and class *Y*? That is, compute *I*(*Y*; *A*).

$$I(Y; A) = H(Y) - H(Y|A) = 1 - 0.68 = 0.32$$

(d) What is the **information gain** between attribute *B* and class *Y*?

$$H(Y|B) = P(B=T)H(Y|B=T) + P(B=F)H(Y|B=F)$$
$$= 50/100[-P(Y=+|B=T)\log P(Y=+|B=T) - P(Y=-|B=T)\log P(Y=-|B=T)]$$
$$+ 50/100[-P(Y=+|B=F)\log P(Y=+|B=F) - P(Y=-|B=F)\log P(Y=-|B=F)]$$
$$= .5[-30/50 \log 30/50 - 20/50 \log 20/50]$$
$$+ .5[-20/50 \log 20/50 - 30/50 \log 30/50]$$

```
   = .5[-(.6)(-.74)–(.4)(-1.32)] + .5[-(.4)(-1.32)–(.6)(-.74)]
   = 0.97
I(Y; B) = 1 – 0.97 = 0.03
```

(e) What is the **information gain** between attribute *C* and class *Y*?

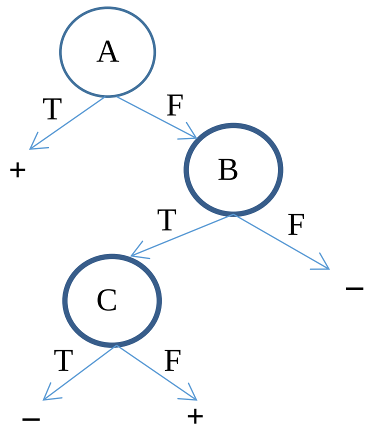
```
H(Y|C) = P(C=T)H(Y|C=T) + P(C=F)H(Y|C=F)
 = 50/100[-P(Y=+|C=T)logP(Y=+|C=T)-P(Y=-|C=T)logP(Y=-|C=T)]
    + 50/100[-P(Y=+|C=F)logP(Y=+|C=F)-P(Y=-|C=F)logP(Y=-|C=F)]
 = .5[-25/50 log 25/50 – 25/50 log 25/50]
    + .5[-25/50 log 25/50 – 25/50 log 25/50]
 = 1.0
I(Y; C) = 1 – 1 = 0.0
```

(f) Manually create the full **decision tree** using the attributes *A*, *B* and *C* to predict the class of *Y*. Show the resulting tree. At each non-leaf node, show the information gain of ALL candidate attributes possible at that node. In case of ties, use the following tie-breaking rules:

   i. For class label majority vote ties, prefer the class “+”.
   ii. For attribute ties, prefer the attribute earliest in the alphabet.

```
Based on (a)-(e), the root node has attribute A.  The child of
the root where A=T is a leaf with class + because all 25
examples are +.  The A=F child has 75 examples; for this node
we have:
H(Y) = -25/75 log 25/75 – 50/75 log 50/75 = 0.91
H(Y|B) = P(B=T)H(Y|B=T) + P(B=F)H(Y|B=F)
   = 45/75[-P(Y=+|B=T)logP(Y=+|B=T)-P(Y=-|B=T)logP(Y=-|B=T)] +
30/75[-P(Y=+|B=F)logP(Y=+|B=F)-P(Y=-|B=F)logP(Y=-|B=F)]
   = .6[-25/45 log 25/45 – 20/45 log 20/45] + .4[-0/30 log
0/30 – 30/30 log 30/30]
   = .593 + 0 = .593
I(Y; B) = .91 - .593 = 0.317
H(Y|C) = P(C=T)H(Y|C=T) + P(C=F)H(Y|C=F)
   = 25/75[-P(Y=+|C=T)logP(Y=+|C=T)-P(Y=-|C=T)logP(Y=-|C=T)] +
50/75[-P(Y=+|C=F)logP(Y=+|C=F)-P(Y=-|C=F)logP(Y=-|C=F)]
   = .33[-0/25 log 0/25 – 25/25 log 25/25] + .67[-25/50 log
25/50 – 25/50 log 25/50]
   = 0 + .67 = .67
I(Y; C) = .91 - .67 = 0.24
So attribute B is selected for the A=F child of the root.
This node’s B=F child has only – examples, so it is a leaf
with class -.  The B=T child is next.  Only attribute C
remains, so C is selected for this node.  Its C=T child has
only – examples, so it is a leaf with class -.  Its C=F child
```

```
has only + examples, so it is a leaf with class +.  So, the
final decision tree is:
```



(g) What is the classification accuracy of your tree on the training set?

```
Because there is no noise in the data, 100% of the training
examples are correctly classified.
```