CSE473 Homework #3
Due at 10:30 AM November 27th in Class on Paper
Name:

Data

For the following problems, use $a_1 \ldots a_4$ as your attributes, $v \in \{0,1\}$ as your target classification, and the following examples as your training set.

Example	a_1	a_2	a_3	a_4	v
1	0	0	1	0	0
2	0	1	0	0	0
3	1	0	0	1	1
4	0	1	1	0	0
5	1	1	0	0	0
6	0	1	0	1	0

Use this example as your test set.

Example	a_1	a_2	a_3	a_4	v
1	0	0	1	1	1

Decision Trees

Root

Show the information gain using each of the attributes to split the examples for the root of the decision tree.

Children

Assume a_2 was chosen as the attribute with the largest information gain. Show the information gains for the $a_2 = 0$ branch of the decision tree.

Hypothesis Space

What does the similarity in calculated information gains from the previous two questions say about the size of the decision tree hypothesis space?

Pruning

Given the classification of the example in the test set, would you prune this tree? Is this test set large enough? What could we do to generate more test set data without access to any more examples?

Naive Bayes

Classifer

Compute the classification for the test value using a naive bayes classifier. Clearly write out the terms for $P(v)$ and $P(a|v)$.

Estimating Probabilities

Now use a m-estimate (see page 179 of mitchell) with the prior estimate of the probability $p = 0.5$ and equivalent sample size of $m = 1$ for the zero probability terms $P(a_1 = 0|v = 1)$ and $P(a_3 = 1|v = 1)$.

Classification Revisted

Now try to classify the instance using m-estimates on just the two terms from the previous problem.

Applicability

Is the answer correct? Comment on the applicability of Naive Bayes to this learning problem.

Expectation Maximization

Harsh expectation maximization is a discrete form of the standard expectation maximization. K-means clustering is an instance of harsh EM. K-means cluster analysis can be used to group N points in a plane into K clusters. Because each point is assigned to just one cluster instead of fractionally to all clusters, it is an example of harsh expectation maximization.
To initialize the algorithm, randomly assign the N points to K clusters. The algorithm repeats the following steps until the assignment of points to clusters converges.

1. Compute the mean of the points assigned to a cluster (the expectation step).

2. Re-assign each point to the cluster whose mean is closer than any other cluster (the maximization step).

This process maps to the expectation maximization algorithm where the missing parameters are the means of the clusters.
Given $K = 2$ and an initial assignment of points $(1, 2)$ and $(5, 6)$ to cluster 1 and points $(2, 2)$ and $(7, 7)$ to cluster 2, show the expectation and maximization steps until the cluster assignment converges.