Homework Week 8

1. Consider the data set

V1	V2	V3	V4	C(lass)
1	-1	-1	1	-1
1	1	1	-1	-1
1	-1	-1	-1	-1
1	1	1	1	-1
-1	-1	1	1	1
-1	1	-1	1	1
-1	1	-1	-1	1
-1	-1	1	-1	1

a) Which variable, Vi, would be selected by a decision tree learner of the type described in the text/lecture as the root of the decision tree? Select one choice and explanation.

a) V1, since its values predict C(lass) with apparent certainty

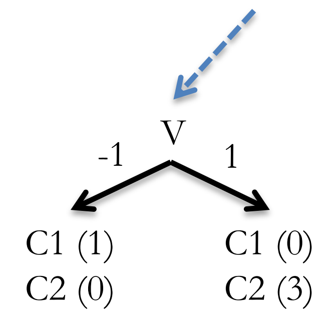
b) V2, since V2 is independent of C(lass)

c) V3, since it lies between the over simplicity of prefect predictiveness and independence

d) V4, since V4 is statistically related to all other variables

e) C(lass), since it is the class

2. Assume that a decision tree has been constructed from training data, and it includes a node that tests on V at the frontier of the tree, with its left child yielding a prediction of class C1 (because the only training datum there is C1), and the right child predicting C2 (because the only training data there are C2). The situation is illustrated here:



Suppose that during subsequent use, it is found that

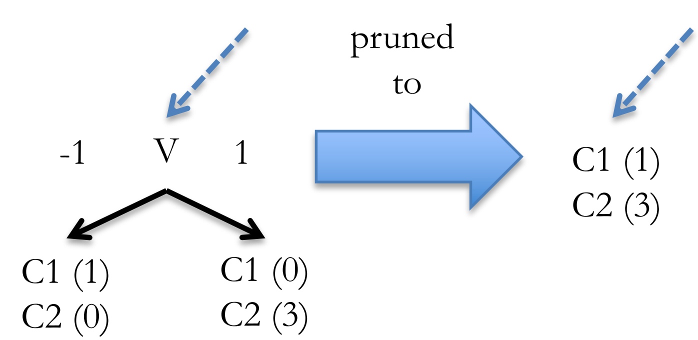
i) a large # of items (N > 1000) are classified to the node with the test on V to the right
ii) 50% of these have V= -1 and 50% of these have V = 1
iii) post classification analysis shows that of the N items reaching the node during usage, 25% were C1 and 75% were C2
iv) of the 0.5 * N items that went to the left leaf during usage, 25% were C1 and 75% were C2
v) of the 0.5 * N items that went to the right leaf during usage, 25% were also C1 and 75% were C2

a) What was the error rate (as a real number rounded to one decimal place) on the sample of N items that went to the sub-tree shown above?

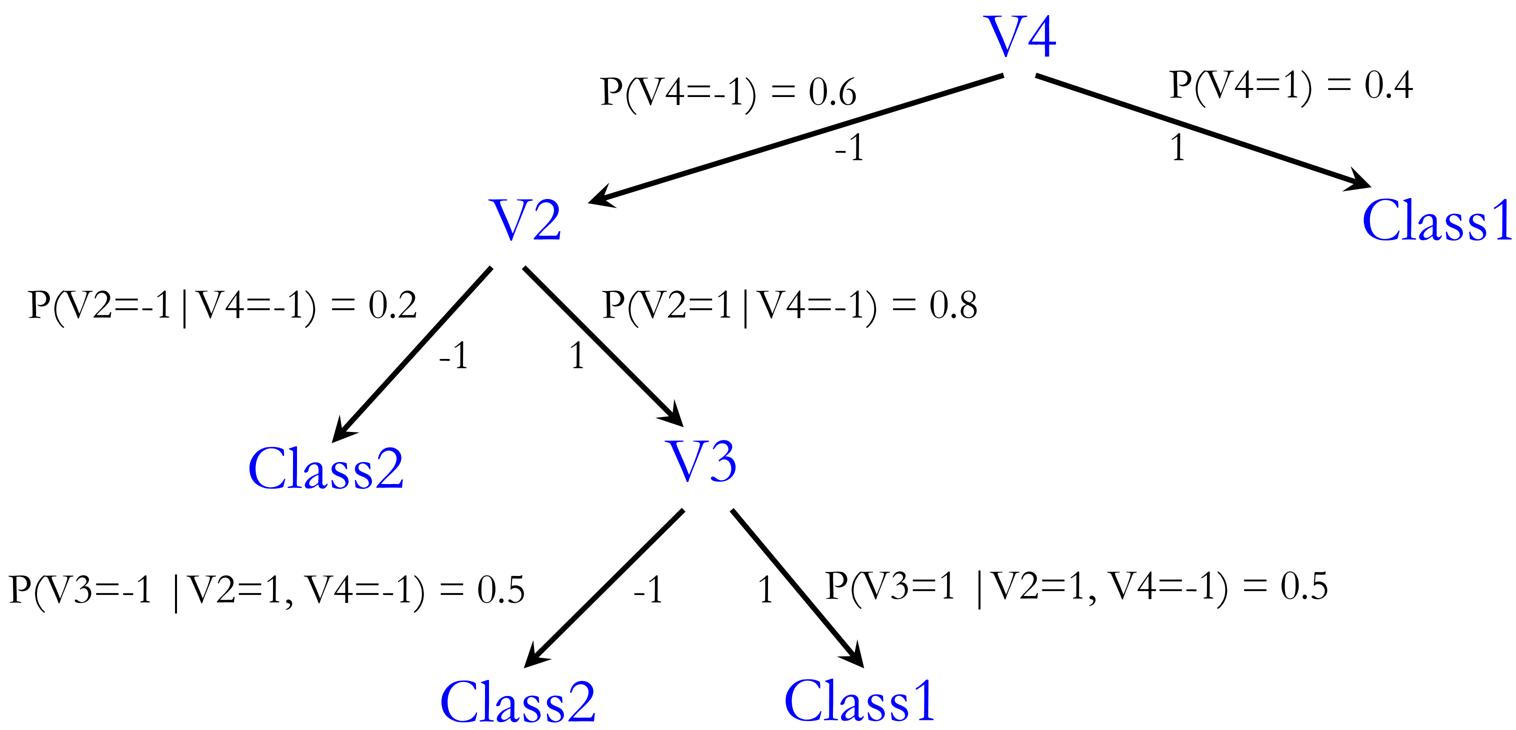
= (.5 * .75) + (.5 * .25)
= .375 + .125
ER = .50

b) What would the error rate (as a real number rounded to two decimal places) on the same sample of N items have been if the sub-tree on previous page (and reproduced here) had been pruned to not include the final test on V, but to rather be a leaf that predicted C2?

= (.5 * .25) + (.5 * .25)
= .125 + .125
ER = .25



3. Consider the following decision tree, where each variable (including the Class variable) is binary valued. Along each branch is the probability the branch will be taken if classification reaches the node from which the branch emanates.



a) Give the expected number of internal nodes visited (i.e., the expected number of variable tests carried out) when classifying an arbitrary datum. **Give the answer as a real number (rounded to two decimal places).**

= 2.08

(calc on next page)

	V4 -> V2	V2 -> V3			levels	
3 levels	0.6	0.8	0.48		3	1.44
2 levels	0.6	0.48	0.12		2	0.24
1 level			0.4		1	0.4
						2.08

3 continued.

b) When classifying a datum with all known values, the classification will proceed along exactly one path of the tree. For example, if the test datum was (V1=1, V2 = -1, V3 = 1, V4 = -1) then classification would proceed down the leftmost path resulting in a Class2 prediction (with probability 1.0). But if there are missing values among the variable values of a test datum, classification may be nondeterministic (or probabilistic).

Suppose that the test datum were (V1=1, V2 = 1, V3 = ?, V4 = ?) – that is, the values of V3 and V4 are not known. What would you conclude is (as a single real number rounded to one decimal place) the probability of a Class1 classification = ?

= .64 or rounded up to 70%
(calc on next page)

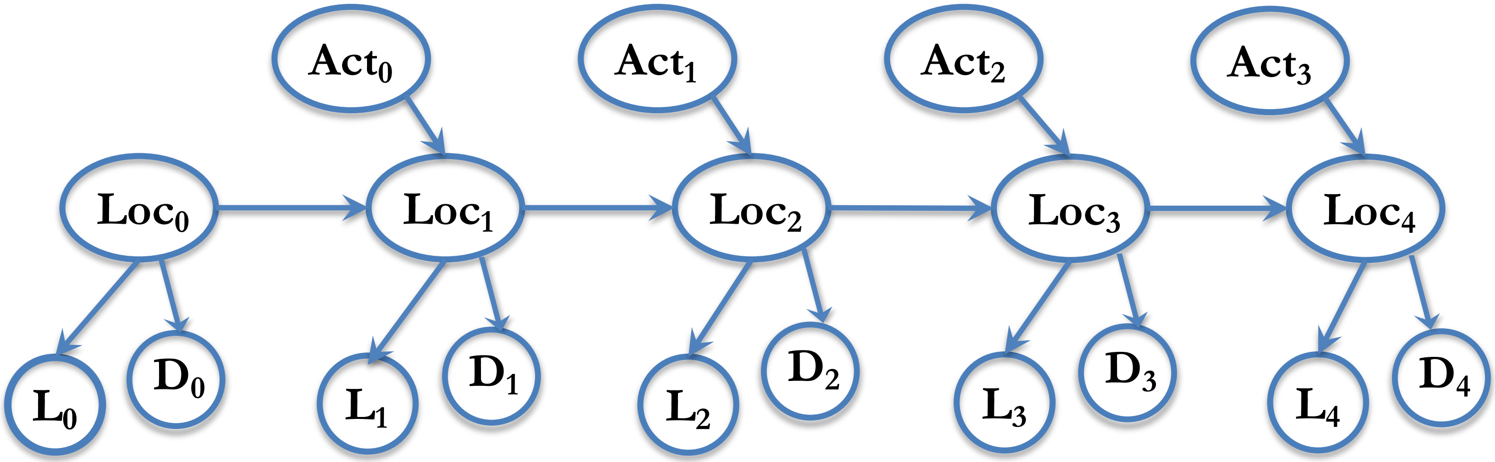
c) Explain how you obtained the result of Part (b)

I calculated all the probabilities of all arcs leading to Class1 for each side of the parent node V4. Then added those probabilities for left and right sides.

#3B						
P(V2=1 | V3=-1)	0.8	P(V3=1 | V2=1, V4=-1)	0.5	=.8 * .5	0.4	
P(V4=1)	0.6	P(V4=-1)	0.4	=.6 * .4	0.24	
					0.64	
					** round up to 70%	

4. Consider this hidden Markov model. Assume all variables are binary valued. Assume probability tables explicitly contain both $P(x|y)$ and $P(\sim x|y)$, so do not need to compute $P(\sim x|y)$.

The domain of Loc_k is $\{c, \sim c\}$, L_i is $\{l, \sim l\}$, D_j is $\{d, \sim d\}$, Act_i is $\{a, \sim a\}$



Assume that the following are observed: $L_0 = \sim l$, $D_0 = d$, $Act_0 = a$, $L_1 = \sim l$, $D_1 = \sim d$

a) What two probabilities would have to be compared to determine which was more probable, $Loc_1 = c$ or $Loc_1 = \sim c$, given the observations above. Ensure that your answer only contains probabilities that would be found in the probability tables for the above HMM.

```
P(Loc0=c) P(Act0=a) P(L0=~l|Loc0=c) P(D0=d|Loc0=c) P(Loc1=c|Loc0=c, Act0=a) P(L1=~l|Loc1=c) P(D1=~d|Loc1=c) +  P(Loc0=~c) P(Act0=a) P(L0=~l|Loc0=~c) P(D0=d|Loc0=~c) P(Loc1=c|Loc0=~c, Act0=a) P(
L1=~l|Loc1=c) P(D1=~d|Loc1=c)


P(Loc0=c) P(Act0=a) P(L0=~l|Loc0=c) P(D0=d|Loc0=c) P(Loc1=~c|Loc0=c, Act0=a) P(L1=~l|Loc1=~c) P(D1=~d|Loc1=~c) + P(Loc0=~c) P(Act0=a) P(L0=~l|Loc0=~c) P(D0=d|Loc0=~c) P(Loc1=~c|Loc0=~c, Act0=a) P(
L1=~l|Loc1=~c) P(D1=~d|Loc1=~c)
```

5. Consider this hidden Markov model. The augmented HMM defines a search space that is qualified by probabilities. Consider the partial search tree below. Give the probabilities of each state for the nodes labeled S_1=c and S_2=f in terms of only probabilities found in the HMM probability tables. Assume that S_0=a is known with certainty and Actions are known with certainty (but their effects are not)