```
1-P(S2|D1,~D2) = (1-Pc(S2|D1,~D2))*(1-Pc(S2|leak))
 Pc(S2|D1,~D2) = 1 - [(1-P(S2|D1,~D2))/(1-Pc(S2|leak)]
               = 1 - [(1 - 0.68)/(1 - 0.20)]
               = 0.60
and,
1-P(S2|~D1,D2) = (1-Pc(S2|~D1,D2))*(1-Pc(S2|leak))
 Pc(S2|~D1,D2) = 1 - [(1-P(S2|~D1,D2))/(1-Pc(S2|leak)]
               = 1 - [(1 - 0.60)/(1 - 0.20)]
               = 0.50

Substituting above gives,
(1-P(S2|D1,D2) = (1-Pc(S2|D1,~D2))(1-Pc(S2|~D1,D2))(1-P(S2|~D1,~D2))
    (1 - 0.84) = (1 - 0.60)*(1 - 0.50)*(1 - 0.20)
          0.16 = 0.16
which is consistent.  Thus the noisy-or assumption is consistent with
this conditional probability table.
```

Problem 3 Learning (20 points)

Part A (2 points)

The closest neighbor is point 7, which is an "eno".

Part B (6 points)

1. The closest 3 neighbors are 7, 8, 9, that is, two "owt" and one "eno" and so the result is "owt".
2. k should be odd to avoid ties.

Part C (8 points)

1.
```
                  Y/ eno
Body Size < 22.5 /
                 \
                  N\ owt
```

2. This predicts "owt".

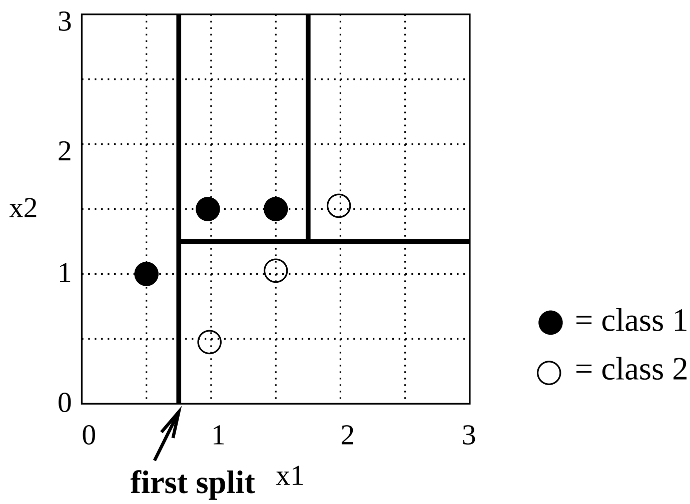
Part D (4 points)

One estimates predictive accuracy by testing the learnt model on "new" data, that is, on data not used to train the model. One standard way is to hold out some of the initial data and use it for training. In this case, there is so little data that this is not possible. So, one is forced to use a technique such as cross-validation in which, for example, one iterates the process of leaving one (or more) of the points out of the training set and then using the rest for training, until all the points have been left out of the training set once. The average of the accuracy of these tests is the predictive accuracy.

Problem 4 Identification Tree (30 Points)

Part A (6 Points)

This is a straightforward test for the use of disorder to build an identification tree. The `make-tree` procedure in PS 8 orders the attribute x_1 before x_2 in the choice of splits in case there are ties.

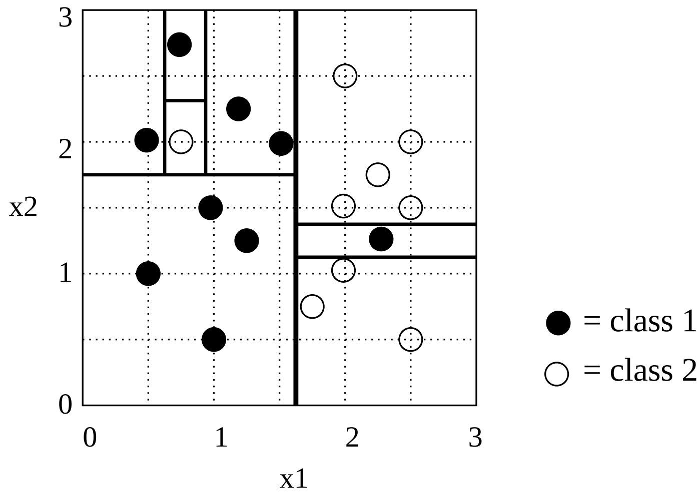


Part B (4 Points)

The top-level split is clear. For the remaining splits, we just look for rectangular regions enclosing the two potential noise points.

Part C (4 Points)

The client is concerned that the excessively fine classifications would decrease the predictive accuracy because of overfitting. The two outliers might just be noise.



Part D (3 Points)

The aggressiveness option gives user some control over the size of the tree. By ignoring the small gain in the reduction of average disorder, the option can produce a 1-level tree for the data set in Part B.

Part E (4 Points)

No. Since there is no reduction in average disorder for any first level split, `make-tree` will not grow any tree.

Part F (5 Points)

E. As N increase, the rectangular decision regions will give a better approximation to the diagonal boundary. However, the improvement will level off as most of the data points will be far away from the decision boundary and hence would not affect the predictive accuracy.

Part G (4 Points)

No. Due to noise points, a deep tree in general overfits the data and hence gives poor predictive accuracy. So we expect the predictive accuracy to increase initially and then decrease afterwards.