1. Assume that you are learning a Boolean function from instances with n Boolean attributes. You are using a hypothesis space that includes conjunctions of Boolean literals.
    a. Define a consistent learning algorithm

Solution: the learning algorithm:

-Begin with h= all literals (positive and negative)

-for each positive instance x take out all literals from h that would make x be classified as negative

Assume x is positive. Then after learning algorithm above h will classify x as positive because it only took out literals (if a hypothesis classifies an instance as positive, then removing more literals from the hypothesis will still classify the instance as positive).

Assume x is negative, then since it is not a positive example the learner would not have removed any literals to make the hypothesis classify it as positive, so the hypothesis will classify it as negative.

    b. Show that such hypotheses are PAC learnable

The algorithm we proposed is polynomial since it just takes time O(n*d) where n is the number of instances and d is the number of attributes. Now we just need to make sure that the sample complexity is polynomial:

Using the bound learned in class the number of samples that we need is at least: $\frac{1}{\epsilon}(\ln(|H|) + \ln\left(\frac{1}{\delta}\right)) = \frac{1}{\epsilon}(\ln(3^d) + \ln\left(\frac{1}{\delta}\right)) = \frac{1}{\epsilon}(\mathrm{d} * \ln(3) + \ln\left(\frac{1}{\delta}\right))$ which is polynomial in all of the parameters.

2. Consider X, the instance space, to be the set of all points in the 2-D plane, i.e. $(x, y) \in R^2$. Give the VC dimension of the following hypothesis spaces:
    a. The set of all rectangles, i.e. $H = \{((a < x < b) \wedge (c < y < d)) | a, b, c, d \in R\}$

Solution VC(H) = 4

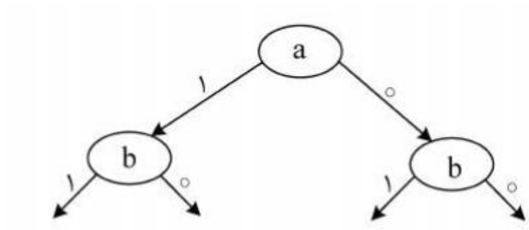
    b. The set of all circles in the 2d plane, where points inside the circle is labeled as positive

Solution: VC(H)=3

    c. Triangles in the 2d plane. Points inside the triangle are classified as positive.

Solution: VC(H)=7

3. Consider a decision tree learning algorithm that considers only examples described by Boolean features $X_1, \ldots, X_n$, learns only Boolean-valued functions (Y ∈ {+, −}), and outputs only 'regular, depth-2 decision trees.' A 'regular, depth-2 decision tree' is a depth two decision tree (a tree with four leaves) in which the left and right child of the root are required to test the same attribute. For example, the tree in picture below is a 'regular, depth-2 decision tree.'



a. Suppose you have noise-free training data for target concept c which you know can be described by a regular, depth-2 decision tree. How many training examples is enough to guarantee that the learner will output a tree whose true accuracy is at least 0.97 with probability 0.99? Assume you have data with 20 attributes in total (although you assume only two of these are needed to describe the correct tree). Also labels of leaves of some attribute can be all positive or all negative- that is you do not have to say that the left child is negative and the right child is positive, they can both be negative or positive.

b. Can you improve this bound? Hint: many trees are actually identical hypothesis functions

Solution: Since |H| is finite, we can use the bound $m \geq \frac{1}{\epsilon}\ (ln(1/\delta)\ +\ ln|H|)$. The number of unique hypotheses is 20 $*$ 19 $*$ 2 4 = 6080. One could plug this in and get a loose bound. However, if we consider all of the hypotheses that are equivalent (many hypotheses provide the same mapping from the instance space to the output space— for example, the ordering of the two attributes that are chosen does not matter), we can find a tighter bound. There are 2 hypotheses that label all examples with one label (either all positive or negative). There are 20 $*$ 2 hypotheses that split on one attribute, and label one side positive, and the other negative. There are $\binom{20}{2}$*10 different hypotheses that split on two attributes and assign labels to the leaves that are not equivalent to the hypotheses mentioned previously (to see this, consider that there are 2^4 = 16 ways to assign labels to the leaves given some choice of attributes to split on, but the cases {++++, - - - -, ++ - -, - - ++, + - + -, - + - +} have already been covered). Thus, |H| = 2 + 20 $*$ 2 + $\binom{20}{2}$ $*$ 10 = 1942. So our bound is: $m \geq \frac{1}{.03}\ (ln(1/.01)\ +\ ln(1942))\ =\ 405.888$. So 406 examples are required.

c. Now consider all regular depth-2 decision trees who can accept real values attributes and split on one threshold. The second level attribute must use the ***same*** threshold for both branches. What is the VC dimension of this hypothesis class?

Solution: 4

4. Determine the VC dimension of where instances are points on the real line and the Hypothesis class is k unions intervals on the real line.

Solution: We can see that we can shatter 2k points- in the worst case every consecutive pair of points is labeled the opposite of its neighbor. For this we only need k intervals to classify all of the positives as positive. So we showed that the VC dimension is at least 2k. To show that it is at most 2k lets take any 2k+1 points. For all points, ordered by value (decreasing to increasing) label each one alternating positive then negative. Since no two positive points can be in the same interval (because there is a negative one in between them) we need at least k+1 intervals. Thus we cannot shatter 2k+1 points.