Phi Duong

2/26/2025

CSE 5693

**Machine Learning Written Assignment 2**

A) 2.4 Consider the instance space consisting of integer points in the x, y plane and the set of hypotheses H consisting of rectangles. More precisely, hypotheses are of the form a <= x <= b, c <= y <= d, where a, b, c, and d can be any integers.

(a) Consider the version space with respect to the set of positive (+) and negative (-) training examples shown below. What is the S boundary of the version space in this case? Write out the hypotheses and draw them in on the diagram.

A graph with numbers and lines

AI-generated content may be incorrect.

The S boundary would be the rectangle where all positive examples are inside the box and all negative examples are left outside the rectangle while minimizing the size of the rectangle. In this case the S boundary would be where <a=4, b=6, c=3, d=5>.

A graph with a square and a line

AI-generated content may be incorrect.

(b) What is the G boundary of this version space? Write out the hypotheses and draw them in

Since the hypothesis includes the boundary, we cannot allow and negative examples to lie along the boundary, so we must take the next smaller integer that maximizes the space. The G boundary would be where <a=3, b=8, c=2, d=7>

A blue square with white lines

AI-generated content may be incorrect.

(c) Suppose the learner may now suggest a new x, y instance and ask the trainer for its classification. Suggest a query guaranteed to reduce the size of the version space, regardless of how the trainer classifies it. Suggest one that will not.

To reduce the size of the version space, we want to query an option (rectangle) that is in between the G and S boundaries. Therefore, any space that consists of 3 <=a <= 4, 6<=b<=8, 2<=c<=3, 5<=d<=7 would be all possible combinations of hypothesis in the version space. As long as we pick a point that satisfies at least one of these hypothesis, we are guaranteed to reduce the version space. An example of this would be <x=5, y =6> A suggestion that would not help would be any examples that were not in the range specified previously. An example of this would be < x =1, y = 1>

(d) Now assume you are a teacher, attempting to teach a particular target concept (e.g., 3 <= x <= 5,2 <= y <= 9). What is the smallest number of training examples you can provide so that the CANDIDATE-ELIMINATION algorithm will perfectly learn the target concept?

To find the minimum, we need to choose examples that reduce the version space as much as possible. To reduce G, we need a negative example. To increase S, we need positive examples. Therefore, we need at least 2 examples (1 positive and 1 negative) for each hypothesis to limit the version space to 1 single target concept, which totals up to **8 examples**

B) 2.7 Consider a concept learning problem in which each instance is a real number, and in which each hypothesis is an interval over the reals. More precisely, each hypothesis in the hypothesis space H is of the form a < x < b, where a and b are any real constants, and x refers to the instance. For example, the hypothesis 4.5 < x < 6.1 classifies instances between 4.5 and 6.1 as positive, and others as negative. Explain informally why there cannot be a maximally specific consistent hypothesis for any set of positive training examples. Suggest a slight modification to the hypothesis representation so that there will be.

We cannot find the maximum specific and consistent hypothesis for this scenario because of the infinite number line. Lets say the lowest example number is x. Therefore, our a hypothesis has to be a number that is not exactly at x but lower than x (if its equal to ex our hypothesis would consider it a negative example). The number line is infinite with “no gaps,” which means that there will always be aa number between our a hypothesis and x. No matter how close a is to x, it will not be the most specific since you can always find an a’ hypothesis that is smaller than the current one. This is also true for b.

A slight modification that could fix this problem is to not include all real numbers but only integers. In this way there can be a number that is 1 unit away from the smallest and highest examples.

C) 3.4ID3 searches for just one consistent hypothesis, whereas the CANDIDATEELIMINATION algorithm finds all consistent hypotheses. Consider the correspondence between these two learning algorithms.

(a) Show the decision tree that would be learned by ID3 assuming it is given the four training examples for the Enjoy Sport? target concept shown in Table 2.1 of Chapter 2.

A diagram of different weather conditions

AI-generated content may be incorrect.

(b) What is the relationship between the learned decision tree and the version space (shown in Figure 2.3 of Chapter 2) that is learned from these same examples? Is the learned tree equivalent to one of the members of the version space?

The relationship between the decision tree and the version space is that, the decision tree creates a hypothesis that is more general than the other hypothesis space.

(c) Add the following training example and compute the new decision tree. This time, show the value of the information gain for each candidate attribute at each step in growing the tree.

Candidates step 1:

InformationGain(S, sky) = Entrop(S) – 4/5 \* Entropy(Sv)

= 0.4421 – 4/5(0.811) = -0.206

InformationGain(S, airTemp) = 0.4421 – 4/5(0.811) = -0.206

InformationGain(S, humidity) = 0.4421 – [2/5(1) + (3/5)0.56366 ] = -0.296

InformationGain(S, wind) = 0.4421- [0 + 0 + (3/5)(0.551) = 0.111

InformationGain(S, water) = 0.4421 – [1/5(0) + 4/5(1)] = -0.3579

InformationGain(S, Forecast) = 0.4421 – [(3/5)(0.9183) + (2/5)(1)] =-0.5089

A diagram of a wind

AI-generated content may be incorrect.

Candidates step 2:

InformationGain(S2,sky) = 0.5282 – 0 = 0.5282

InformationGain(S2, airTemp) = 0.5282

InformationGain(S2, Humidity) = 0.5282 – (1/4(0) + ¾(0.9183)) = -0.1605

InformationGain(S2, Water) = 0.5282 – (1/4(0) + ¾(0.9183)) = -0.1605

InformationGain(S2, Forecast) = 0.5282 – [2/4(0) + 2/4(1)] = 0.0282

A diagram of a weather forecast

AI-generated content may be incorrect.

(d) Suppose we wish to design a learner that (like ID3) searches a space of decision tree hypotheses and (like CANDIDATE-ELIMINATION) finds all hypotheses consistent with the data. In short, we wish to apply the CANDIDATE-ELIMINATION algorithm to searching the space of decision tree hypotheses. Show the S and G sets that result from the first training example from Table 2.1. Note S must contain the most specific decision trees consistent with the data, whereas G must contain the most general. Show how the S and G sets are refined by the second training example (you may omit syntactically distinct trees that describe the same concept). What difficulties do you foresee in applying CANDIDATE-ELIMINATION to a decision tree hypothesis space?

G trees:

A diagram of different weather conditions

AI-generated content may be incorrect.

S tree:

A diagram of a weather forecast

AI-generated content may be incorrect.

To make a s tree hypothesis, we need to find a tree that has the most levels. Therefore, the most specific trees are the ones that go down on branch. Unlike most decision trees where it is roughly evenly distributed on both sides this tree is branching off one side. This isn’t ideal for decision trees. In addition, to find the most specific, you must look at all three combinations since it’s not easy to rearrange any given tree. It is also difficult to follow the find s algorithm since you must start at the most specific, which is hard to determine with decision trees

D) Consider two attributes Outlook (sunny, rainy, cloudy) and Humidity (high) and outcome PlayTennis (yes, no) for the instance space (X).

1. Consider an unbiased hypothesis space (H1), enumerate all possible hypotheses (h1, h2, ...) in terms of subsets of instances. What is the number of possible unique hypotheses in H1?

2^(5\*3) = 32768

1. For each hypothesis in H1, represent it as a boolean expression. What is the number of unique hypotheses semantically?

2^(4\*2) + 1 = 257

1. Consider a biased hypothesis space (H2) where each attribute can only have a value, ?, or ∅. What is the number of unique hypotheses semantically in the biased hypothesis space (H2)?

2^2 = 4

1. Identify hypotheses in the unbiased hypothesis space (H1) that are not in the biased hypothesis space (H2).

Hypothesis space in H1 that is not in h2 is any specific hypothesis other than ? or 0, therefore <sunny, ?> is not found in the hypothesis space

(e) With the programming assignment: Discuss and compare accuracy of no pruning versus rule post-pruning in testIris and testIrisNoisy. Include plots for the comparisons.

A graph of a graph with red and blue bars

AI-generated content may be incorrect.

Table1: shows the accuracy of no pruning vs rule post pruning based on training set and test set.

A graph with red and blue lines

AI-generated content may be incorrect.

Table2: shows the accuracy of no pruning vs rule post pruning as percent noise increases

Table 1 shows the increase in accuracy of the test set when using rule post-pruning at the cost of reducing accuracy for training cases. This is more ideal since we want to generalize the machine learning model to be capable of handling new instances better. Table 2 shows this even better since the accuracy for rule post-pruning stayed consistent as the percentage noise increased in the training set. In addition, the no prune model lost some accuracy due to the noise messing up the tree model. This shows that pruning keeps a high accuracy even when there are errors in the training set.