**CSE 5693 Machine Learning**

**HW2 Decision Tree Learning**

Josias Moukpe

Written Assignment

1. **2.4**

**Instance space consist of integer points in x, y plane and H is the set of hypotheses consisting of rectangles. The hypotheses are of the form a ≤ x ≤ b, c ≤ y ≤ d where a, b, c, and d are integers.**

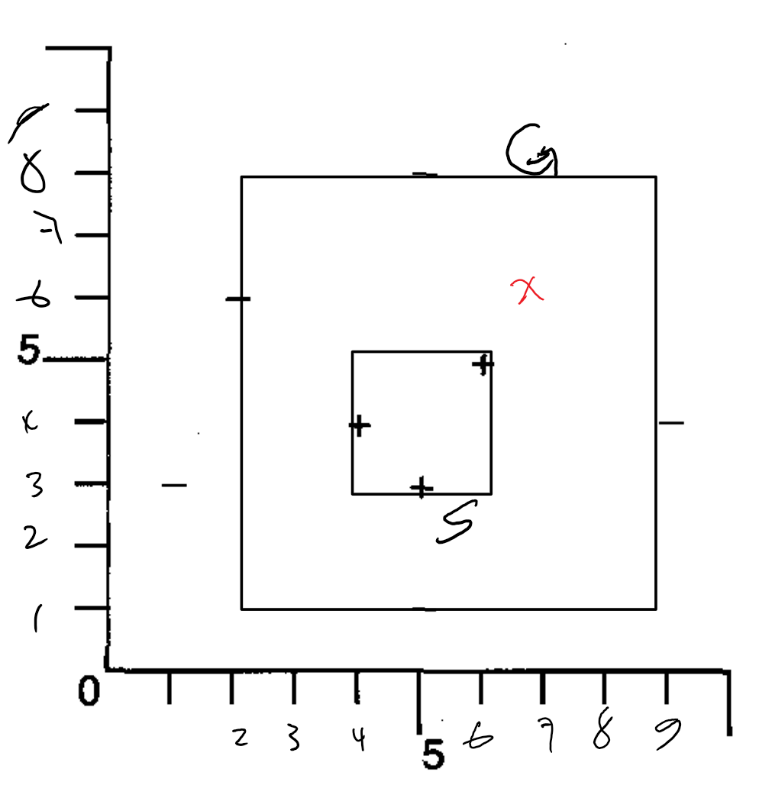
****

Figure 1. Drawing of S and G boundaries

* 1. **What is the S boundary of the version space?**

S: 4 ≤ x ≤ 6, 3 ≤ y ≤ 5

See figure 1 for drawing of S

* 1. **What is the G boundary of the version space?**

G: 2 ≤ x ≤ 9, 1 ≤ y ≤ 8

See figure 1. for drawing of G

* 1. **Suggest x, y instance**

P: x = 7, y = 6 is guaranteed to reduce the size of the version space

if P is positive (+), S: 4 ≤ x ≤ 7, 3 ≤ y ≤ 6

if P is negative (-), G: 2 ≤ x ≤ 7, 1 ≤ y ≤ 6

Q: x = 2, y = 1 is not guaranteed to reduce the size of the version space

Since if Q is negative (-), G doesn’t change

* 1. **Teacher**

A minimum, you need 4 training examples, since the rectangle can be described by 2 pairs of points, one pair of positive points and another pair of negative points to set the S and G limits. For example, positive pair {(3,2), (5, 9)} and negative pair {(2,1), (6, 10)} is enough for candidate eliminate to learn target: 3 ≤ x ≤ 5 and 2 ≤ y ≤ 9

1. **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.**

Instances are real numbers and a hypothesis is of the form a < x < b.

Between 2 real numbers, there is an infinity of real numbers so there cannot be any maximally specific hypothesis for positives. A modification of the hypothesis representation would be to use the form a ≤ x ≤ b so the maximally specific hypothesis for positives would be when a = b so a ≤ x ≤ a or b ≤ x ≤ b.

1. **3.4**

**ID3 searches for just one consistent hypothesis, whereas the CANDIDATE-ELIMINATION algorithm finds all consistent hypotheses. Consider the correspondence between these two learning algorithms.**

* 1. **Show the decision tree that would be learned by ID3 assuming it is given the four training examples for the Enjoy Sport?**
  2. **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?**
  3. **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.**
  4. **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 these constraining 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?**

1. **Play Tennis**

**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. **For each hypothesis in H1, represent it as a boolean expression. What is the number of unique hypotheses semantically?**
  3. **Consider a biased hypothesis space (H2) where each attribute can only have a value, ?, or null. What is the number of unique hypotheses semantically in the biased hypothesis space (H2)?**
  4. **Identify hypotheses in the unbiased hypothesis space (H1) that are not in the biased hypothesis space (H2).**

1. **Discuss and compare accuracy of no pruning versus rule post-pruning in testIris and testIrisNoisy. Include plots for the comparisons.**

**30% is best for validation set**

Figure 3. Comparison of pruning vs no pruning based on noise level

Table 1.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Noise Level (%) | No Pruning Accuracy (%) | | | | | Pruning Accuracy (%) | | | | |
|  | Trial 1 | Trial 2 | Trial 3 | Trial 4 | **Average** | Trial 1 | Trial 2 | Trial 3 | Trial 4 | **Average** |
| 0 | 96 | 96 | 96 | 96 | **96** | 96 | 96 | 96 | 96 | **96** |
| 2 | 96 | 96 | 96 | 96 | **96** | 96 | 96 | 96 | 96 | **96** |
| 4 | 92 | 94 | 88 | 76 | **87.5** | 96 | 96 | 88 | 96 | **94** |
| 6 | 84 | 92 | 94 | 92 | **90.5** | 90 | 96 | 94 | 96 | **94** |
| 8 | 74 | 70 | 86 | 86 | **79** | 90 | 78 | 96 | 96 | **90** |
| 10 | 60 | 64 | 80 | 58 | **65.5** | 90 | 76 | 90 | 96 | **88** |
| 12 | 42 | 66 | 70 | 68 | **61.5** | 54 | 90 | 92 | 96 | **83** |
| 14 | 22 | 54 | 40 | 46 | **40.5** | 30 | 64 | 64 | 86 | **61** |
| 16 | 36 | 52 | 54 | 62 | **51** | 48 | 46 | 76 | 50 | **55** |
| 18 | 46 | 30 | 56 | 64 | **49** | 50 | 60 | 88 | 48 | **61.5** |
| 20 | 50 | 32 | 40 | 38 | **40** | 54 | 62 | 32 | 58 | **51.5** |