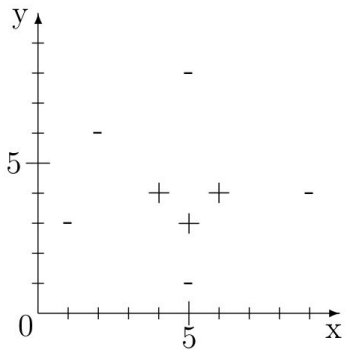
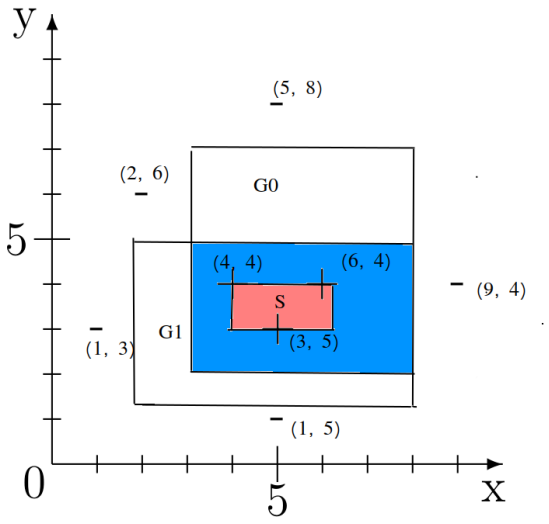
Assignment 4.1

For this exercise, a hypothesis space as shown on slide 33 (concept learning) should be considered. The instance space consists of integer points in the x-y-plane, the hypothesis space are rectangles. More precisely, hypotheses are of the form $a \leq x \leq b,\ c \leq y \leq d$ where $a, b, c, d \in \mathbb{N}$.

a) Consider the version space with respect to the set of positive and negative training examples shown below. What are the S and G boundaries for this case? Write down the hypotheses and draw them in the diagram.



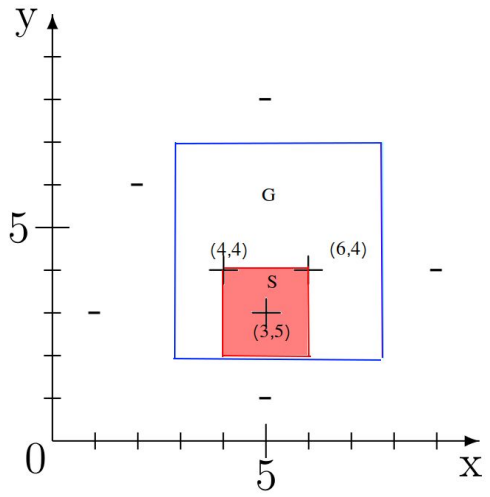


$S : \{4 \leq x \leq 6,\ 3 \leq y \leq 4\}$

$G_0 : \{3 \leq x \leq 8,\ 2 \leq y \leq 7\}$

$G_1 : \{2 \leq x \leq 8,\ 2 \leq y \leq 5\}$

b) Consider the diagram from subtask a) again, but this time use squares as hypotheses. What will change?

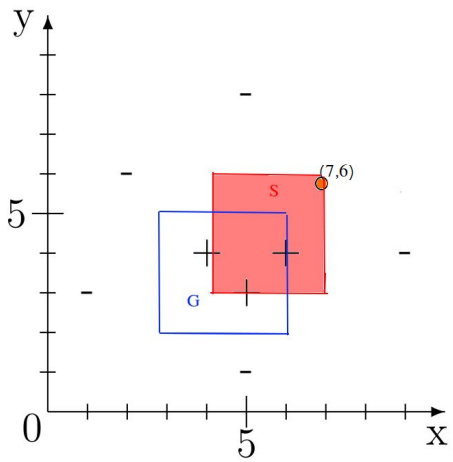


S : {4 ≤ x ≤ 6, 3 ≤ y ≤ 5}

G : {3 ≤ x ≤ 8, 2 ≤ y ≤ 7}

c) Suppose the learner may now suggest a new instance and asks 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, if possible. Consider both cases from a and b.

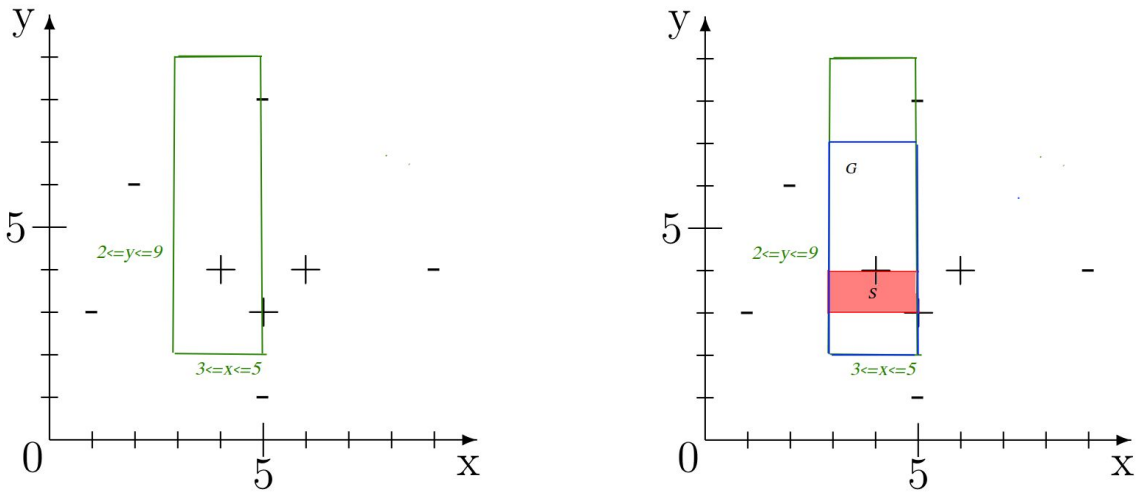
Consider a query point (7,6)



S : {4 ≤ x ≤ 7, 3 ≤ y ≤ 6}

G : {3 ≤ x ≤ 6, 2 ≤ y ≤ 5}

d) Assume that the teacher is attempting to teach a particular target concept (e.g. $3 \leq x \leq 5,\ 2 \leq y \leq 9$). What is the smallest number of training examples you can provide so that the Candidate-Elimination-Algorithm will perfectly learn the target concept? Demonstrate this with the example.



Target Concept: Defines a positive training sample (4, 4)
Instance Space: Points in integer coordinates in 2-D space.
Hypotheses Space: Rectangles; i.e the hypotheses are of the form: $3 \leq x \leq 5,\ 2 \leq y \leq 9$

The target concept is exactly learned when the S and G boundary sets converge to a single, identical, hypothesis.

The smallest number of training example for the CE algorithm
to perfectly learn the target concept = 1 < (4, 4), + >