## Concept Learning

1. 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.
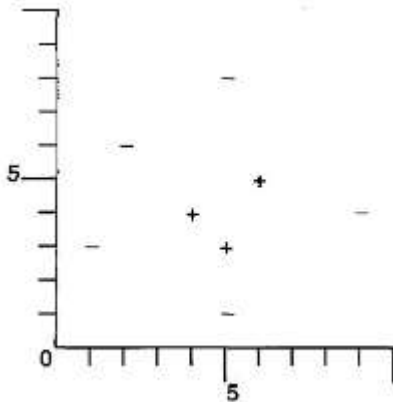
b) What is the G boundary of this version space? Write out the hypotheses and draw them in.

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.

d) Now assume you are a teacher, 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?

2. Consider the following sequence of positive and negative training examples describing the concept "pairs of people who live in the same house." Each training example describes an ordered pair of people, with each person described by their sex, hair color (black, brown, or blonde), height (tall, medium, or short), and nationality (US,French, German, Irish, Indian, Japanese, or Portuguese)

+	$\langle\langle male\ brown\ tall\ US\rangle\langle female\ black\ short\ US\rangle\rangle$
+	$\langle\langle male\ brown\ short\ French\rangle\langle female\ black\ short\ US\rangle\rangle$
−	$\langle\langle female\ brown\ tall\ German\rangle\langle female\ black\ short\ Indian\rangle\rangle$
+	$\langle\langle male\ brown\ tall\ Irish\rangle\langle female\ brown\ short\ Irish\rangle\rangle$

Consider a hypothesis space defined over these instances, in which each hypothesis is represented by a pair of 4-tuples, and where each attribute constraint may be a specific value, "?," or "Φ," just as in the EnjoySport hypothesis representation.

For example, the hypothesis

$$\langle\langle male\ ?\ tall\ ?\rangle\langle female\ ?\ ?\ Japanese\rangle\rangle$$

a) Provide a hand trace of the CANDIDATE-ELIMINATION algorithm learning from the above training examples and hypothesis language. In particular, show the specific and general boundaries of the version space after it has processed the first training example, then the second training example, etc.