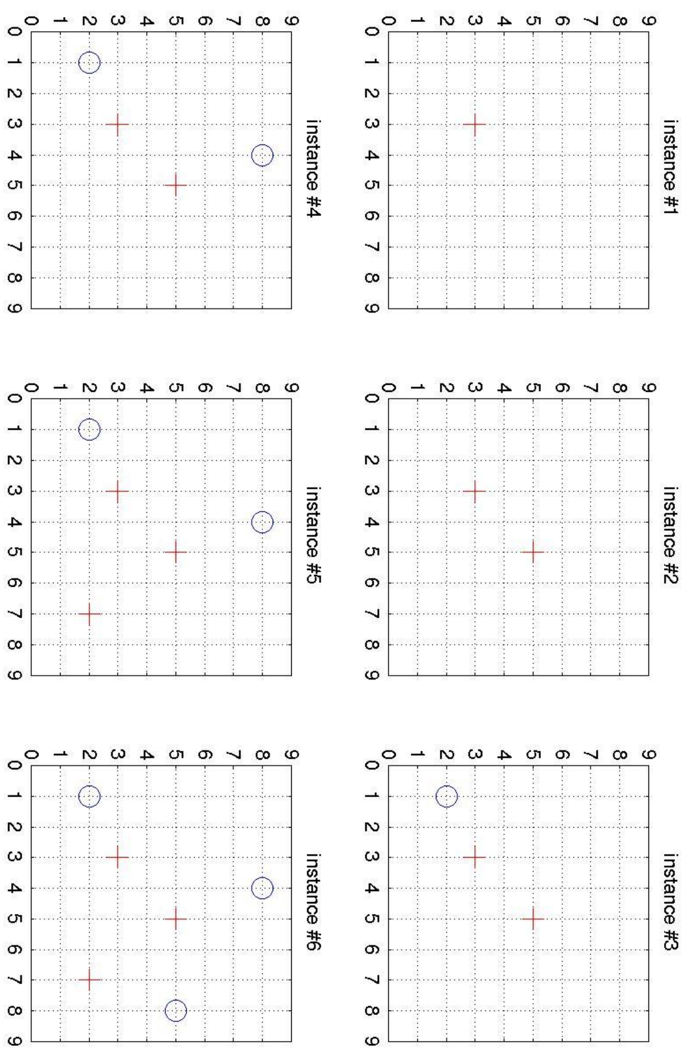
Practice Problems on Concept Learning:

1. Consider the instance space X consisting of integer points in the x,y plane ($0 \le x \le 9$, $0 \le y \le 9$) and the set of hypotheses H consisting of axes-parallel rectangles. More precisely, hypotheses are of the form ($a \le x \le b$, $c \le y \le d$), where a, b, c and d can be integers.

I. Consider the version spaces with respect to the set of positive (+) and negative (o) training examples shown on the second page. Trace the S- and G -boundaries of the version space using the CANDIDATE-ELIMINATION algorithm for each new training instance. Write out the hypotheses that belong to the S- and G-boundary and draw them into the diagram.

II. 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.

III. Now assume that you are a teacher, attempting to teach a particular target concept (e.g. $3 \le x \le 5$, $2 \le y \le 7$). What is the smallest number of training examples you can provide so that the CANDIDATE-ELIMINATION algorithm will perfectly learn the target concept?



2. Suppose you are a doctor treating a patient who occasionally suffers from an allergic reaction. Your intuition tells you that the allergy is a direct consequence of certain foods the patient eat, place where it is eaten, time of day, day of week and the amount spent on food. So we gathered some data and the data is summarized in the table below.

Number	**Restaurant**	**Meal**	**Day**	**Cost**	**Reaction**
1	Sam's	breakfast	Friday	cheap	yes
2	Hilton	lunch	Friday	expensive	no
3	Sam's	lunch	Saturday	cheap	yes
4	Denny's	breakfast	Sunday	cheap	no
5	Sam's	breakfast	Sunday	expensive	no

What is the hypothesis that fits this data using Find-S algorithm. Show trace.

3. Consider the following set of attributes:

Study*: Intense, Moderate, None,* ***Difficulty****: Easy, Hard,* ***Sleepy****: Very, Somewhat,* ***Attendance****: Frequent, Rare,* ***Hungry****: Yes, No,* ***Thirsty****: Yes, No,* ***PassTest****: Yes, No*

Suppose we have the data set:

Example	Study	Difficulty	Sleepy	Attendance	Hungry	Thirsty	Earn-A
1	Intense	Normal	Extremely	Frequent	No	No	**Yes**
2	Intense	Normal	Slightly	Frequent	No	No	**Yes**
3	None	High	Slightly	Frequent	No	Yes	**No**
4	Intense	Normal	Slightly	Frequent	Yes	Yes	**Yes**

Consider the space *H* of conjunctive hypotheses, which, for each attribute, either:

- indicates by a "?" that any value is acceptable for this attribute,
- specifies a single required value (e.g., *Normal*) for this attribute, or
- indicates by a "∅" that no value is acceptable.

Let a version space (a subset of consistent hypotheses in *H*) be represented by an *S* set (specific boundary, at the top) and a *G* set (general boundary, at the bottom). Suppose the 4 training examples above are presented in order.

a) Draw a diagram showing the evolution of the version space for concept ***Earn-A*** given the training examples, by writing down S_1, G_1, S_2, G_2, S_3, G_3, S_4, and G_4. If the *G* set does not change given a new example, just write $G_{i+1} = G_i$ next to the drawing of G_i (similarly for *S*).

b) Write down the hypotheses in the final version space (the ones that lie between S_4 and G_4 according to the partial ordering relation *Less-Specific-Than*).

c) In the final version space, draw lines between hypotheses that are related by this relation. For example, there should be a line between <?, *Normal*, ?, ?, ?> and <?, *Normal*, ?, *Frequent*, ?, ?>.