پاسخ تکلیف مبحث Concept Learning

درس یادگیری ماشین

امیرحسین ابوالحسنی
۴۰۰۴۰۵۰۰۳

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 $a \leq x \leq b$, $c \leq y \leq d$, where $a, b, c,$ and d can be any integers.

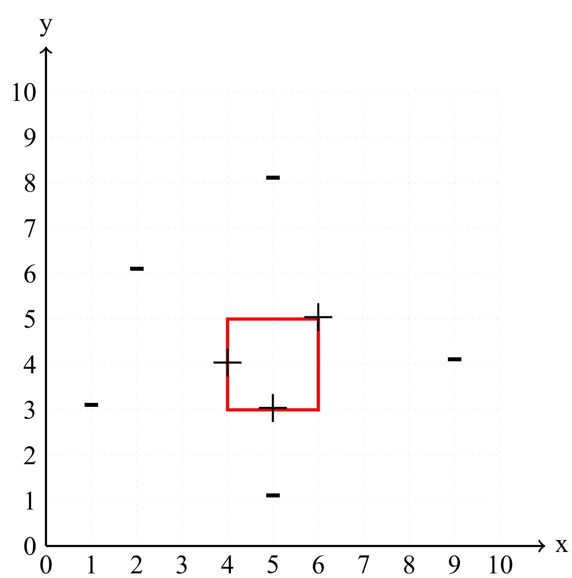
1.1 Consider the version space with respect to the set of positive (+) and negative (−) training examples shown below. What is the S boundary of the version space in this case? Write out the hypotheses and draw them in on the diagram.

Answer:

$$S = \{h\}$$

$$h : 4 \leq x \leq 6 \ \ , \ \ 3 \leq y \leq 5$$

The *red* rectangle is the S boundary :



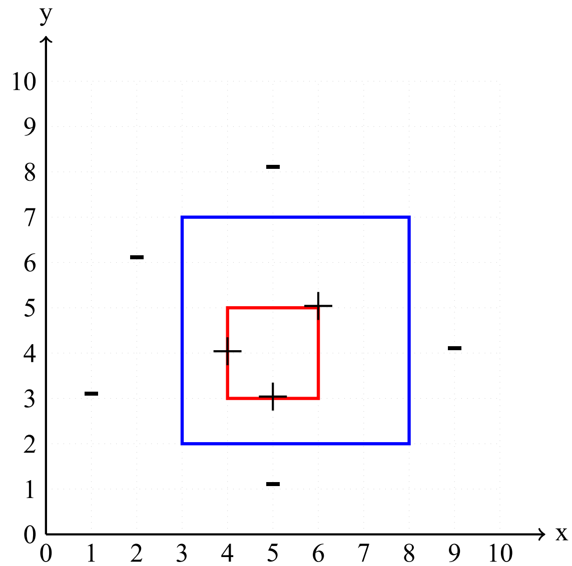
1.2 What is the G boundary of this version space? Write out the hypotheses and draw them in.

Answer:

$$G = \{h\}$$

$$h : 3 \le x \le 8 \ , \ 2 \le y \le 7$$

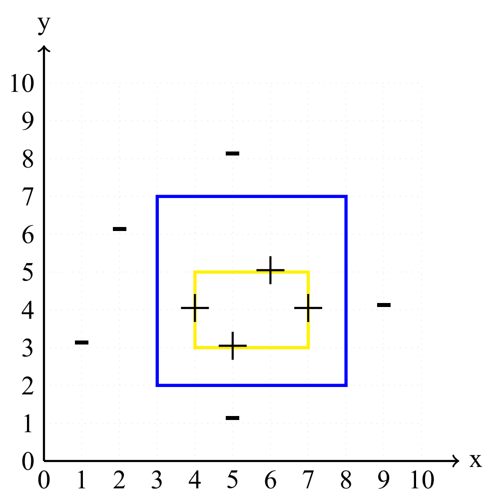
The *blue* rectangle is the G boundary :



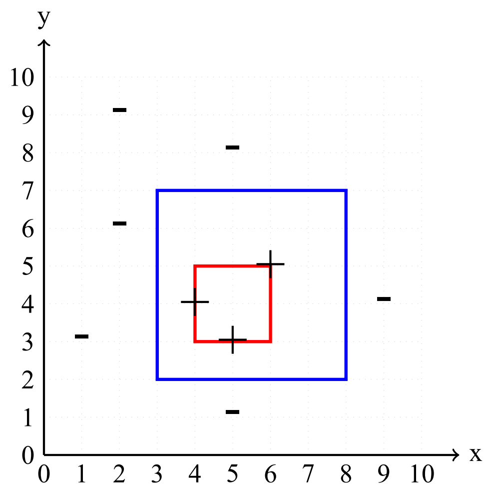
1.3 Suppose the learner may now suggest a new x , y instance and ask the trainer for its classification. Suggest a query guaranteed to reduce the size of the version space, regardless of how the trainer classifies it. Suggest one that will not.

Answer:

If $P = (7, 4)$ and +, then the S boundary will get larger and thus, the size of the version space will get smaller.



If point P is located outside G boundary and is − (or is + and inside the S boundary), it will not cause any changes to the size of the version space. e.g. $P = (2, 9)$



1.4 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?

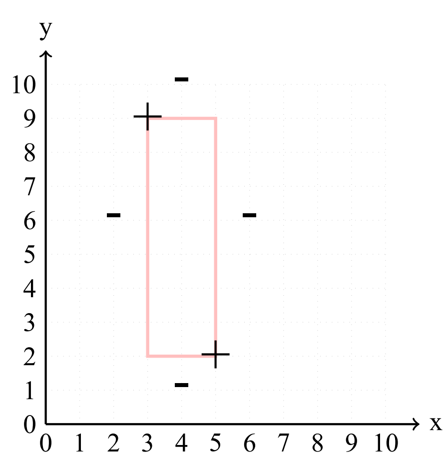
Answer:

I believe we need minimally 6 instances, 2 positive and 4 negative examples learn any hypothesis h **perfectly**.
In order to perfectly learn any hypothesis in this space:

$$\text{Version Space} = \{h\}$$

and for this to happen, G must be equal to S:

$$G = S$$



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)

$$+\langle\langle \textbf{male brown tall US}\rangle\langle\textbf{female black short US}\rangle\rangle$$

$$+\langle\langle \textbf{male brown short French}\rangle\langle\textbf{female black short US}\rangle\rangle$$

$$-\langle\langle \textbf{female brown tall German}\rangle\langle\textbf{female black short Indian}\rangle\rangle$$

$$+\langle\langle \textbf{male brown tall Irish}\rangle\langle\textbf{female brown short Irish}\rangle\rangle$$

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 ”∅” just as in the EnjoySport hypothesis representation.

2.1 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.

Answer:

Initialization:

$$G_0 = \{\ \langle\langle ?, ?, ?, ?\rangle\langle ?, ?, ?, ?\rangle\rangle\}$$

$$S_0 = \{\ \langle\langle \varnothing, \varnothing, \varnothing, \varnothing\rangle\langle \varnothing, \varnothing, \varnothing, \varnothing\rangle\rangle\}$$

Sample 1:

$$G_1 = G_0$$

$$S_1 = \{\ \langle\langle \text{male, brown, tall, US}\rangle\langle\text{female, black, short, US}\rangle\rangle\}$$

Sample 2:

$$G_2 = G_1$$

$$S_2 = \{\ \langle\langle \text{male, brown, ?, ?}\rangle\langle\text{female, black, short, US}\rangle\rangle\}$$

Sample 3:

$$G_3 = \{\ \langle\langle \text{male, ?, ?, ?}\rangle\langle ?, ?, ?, ?\rangle\rangle,\ \langle\langle ?, ?, ?, ?\rangle\langle ?, ?, ?, \text{US}\rangle\rangle\}$$

$$S_3 = S_2$$

Sample 4:

$$G_3 = \{\ \langle\langle \text{male, ?, ?, ?}\rangle\langle ?, ?, ?, ?\rangle\rangle\}$$

$$S_3 = \{\ \langle\langle \text{male, brown, ?, ?}\rangle\langle\text{female, ?, short, ?}\rangle\rangle\}$$