Упражнение 1

Денис Симеонов Михайлов
ФН: 25788

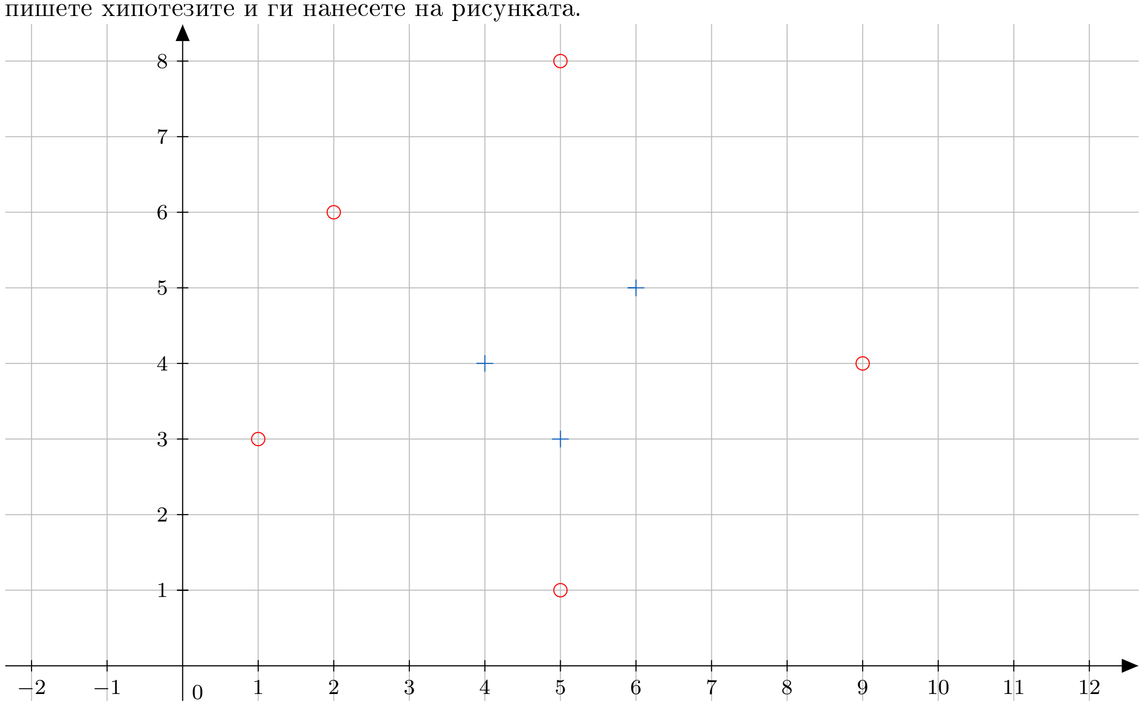
23 октомври 2017г.

Задача Представете си, че пространството на примери се състои от точки $< x, y >$ с целочислени координати, а пространството на хипотези H се състои от правоъгълници. По-точно, хипотезите се записват във вида

$$a \leq x \leq b, c \leq y \leq d$$

където a, b, c и d са цели числа.

1. Разгледайте пространството на версиите по отношение на положителните (+) и отрицателните (o) обучаващи примери, показани по-долу. Коя е S границата на пространството на версиите в този случай? Напишете хипотезите и ги нанесете на рисунката.



2. Коя е G границата на това пространство на версиите? Напишете хипотезите и ги нанесете на рисунката.

Решение Ще използваме алгоритъма за елиминиране на кандидати.

Първо ще дефинираме релацията „по-обща или равна на" за хипотезите от H. Ще казваме, че h_1 от вида:

$$a_1 \leq x \leq b_1, c_1 \leq y \leq d_1$$

е „по-обща или равна на" h_2 от вида:

$$a_2 \leq x \leq b_2, c_2 \leq y \leq d_2$$

ако е изпълнено:

$$\begin{array}{l} a_1 \leq a_2 \\ b_1 \geq b_2 \\ c_1 \leq c_2 \\ d_1 \geq d_2 \end{array}$$

Бележим с $h_1 \geq h_2$

Построяваме множеството с най-общите хипотези в H:

$$G = \{-\infty \leq x \leq \infty, -\infty \leq y \leq \infty\}$$

В G попада единствено хипотезата, което обхваща всички точки в пространството

Построяваме и множеството S от най-специфичните хипотези в H:

$$S = \{\infty \leq x \leq -\infty, \infty \leq y \leq -\infty\}$$

В S попада възможно най-специфичната хипотеза от H. Всяка друга хипотеза от H е „по-обща от" нея.

Нека първо разгледаме положителните примери. При настъпването на обучаващия пример $< 5, 3 >$ G не се променя, защото хипотезата в G е съвместима с всички точки в пространството. Хипотезата в S не е съвместима с нито една точка от пространството и нейното най-малко обобщение, което е съвместима с $< 5, 3 >$ е

$$5 \leq x \leq 5, 3 \leq y \leq 3$$

Хипотезата от G е „по-обща или равна на" нея, защото тя е по-обща от всяка друга хипотеза. Така множеството S вече има вида:

$$S = \{5 \leq x \leq 5, 3 \leq y \leq 3\}$$

Нека следващият обучителен пример да е $< 4, 4 >$. G отново не се променя, а хипотезата в S е несъвместима с тази точка. Това означава, че тя трябва да бъде премахната, а на нейно място да се добави най-малкото и обощение, такова че да е съвместимо с $< 4, 4 >$. Това най-малко обощение е

$$4 \leq x \leq 5, 3 \leq y \leq 4$$

Това, което направихме, беше да увеличим интервала $[5, 5]$ за x до най-малкия възможен, който включва 4. Това е интервалът $[4, 5]$. Аналогично постъпваме и за y. По този начин S придобива вида:

$$S = \{4 \leq x \leq 5, 3 \leq y \leq 4\}$$

След още една аналогична итерация за точката $< 6, 5 >$, S има вида:

$$S = \{h_s = 4 \leq x \leq 6, 3 \leq y \leq 5\}$$

Нека сега видим какво се случва когато обработваме отрицателните обучителни примери. Нека първо постъпи примера $< 5, 1 >$.Хипотезата от G е несъвместима с този пример. Затова трябва да бъде изтрита и на нейно място да се появят нейните най-малки специализации. Това са хипотезите:

$$\begin{aligned}
&h_1 = -\infty \leq x \leq 4, -\infty \leq y \leq \infty \\
&h_2 = 6 \leq x \leq \infty, -\infty \leq y \leq \infty \\
&h_3 = -\infty \leq x \leq \infty, 2 \leq y \leq \infty \\
&h_4 = -\infty \leq x \leq \infty, -\infty \leq y \leq 0
\end{aligned}$$

От тези четири хипотези само $h_3 \geq h_s$ и затова само тя ще бъде добавена към G. По този начин G е придобива вида:

$$G = \{h_g = -\infty \leq x \leq \infty, 2 \leq y \leq \infty\}$$

Нека следващият обучителен пример да е $< 1, 3 >$. Той също е несъвместим с хипотезата h_g. Аналогично на предния обучаващ пример, хипотезата h_g има следните най-малки специализации:

$$\begin{aligned}
&h_1 = -\infty \leq x \leq 0, 2 \leq y \leq \infty \\
&h_2 = 2 \leq x \leq \infty, 2 \leq y \leq \infty \\
&h_3 = -\infty \leq x \leq \infty, 2 \leq y \leq 2 \\
&h_4 = -\infty \leq x \leq \infty, 4 \leq y \leq \infty
\end{aligned}$$

От тези само $h_2 \geq h_s$ и затова тя ще бъде добавена към G. Вида на G е:

$$G = \{h_g = 2 \leq x \leq \infty, 2 \leq y \leq \infty\}$$

Следващият обучителен пример е $< 2, 6 >$. h_g отново е несъвместима с него и нейната най-малката специализация на h_g, такава че h_s да е по-специфична от нея, е:

$$2 \leq x \leq \infty, 2 \leq y \leq 5$$

По този начин G придобива вида:

$$G = \{h_g = 2 \leq x \leq \infty, 2 \leq y \leq 5\}$$

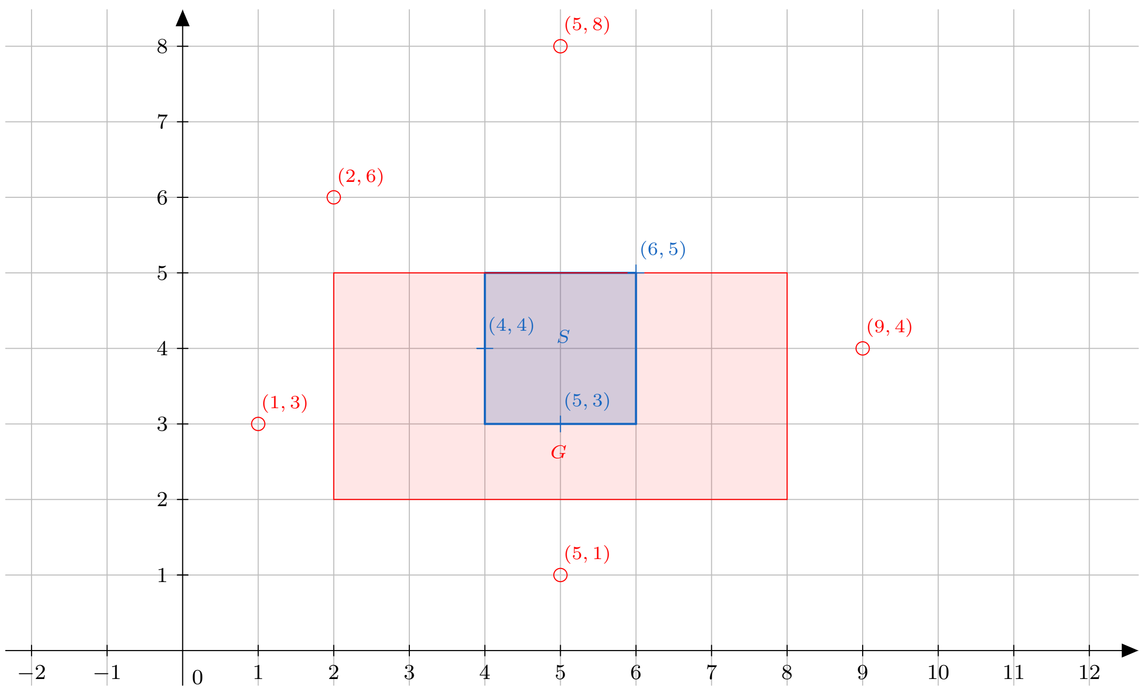
Следващият пример е $< 5, 8 >$, но h_g и h_s са съвместими с него затова нищо не се променя. Последният пример е $< 9, 4 >$. h_g е несъвместима с него и нейната най-малка специализация е:

$$2 \leq x \leq 8, 2 \leq y \leq 5$$

И така крайният вид на множествата е следния:

$$G = \{h_g = 2 \leq x \leq 8, 2 \leq y \leq 5\}$$

$$S = \{h_s = 4 \leq x \leq 6, 3 \leq y \leq 5\}$$



3. Да предположим, че вие трябва да предложите нов пример $< x, y >$ и да запитате учителя за неговата класификация. Предложете заявката, която гарантирано ще намали пространството на версиите независимо от това, как учителят ще й класифицира. Предложете и друга заявка, която няма да намали това пространство.

Решение Ако искаме да намалим пространството на версиите, тогава h_g трябва да е съвместима с примера d, а h_s трябва да е несъвместима с примера d или обратното. И в двата случая това означава,

че примерът d трябва да е извън правоъгълника обособен от S и да е вътре в правоъгълника, обособен от G. Така както и да бъде класифициран примера от учителя, той винага ще е несъвместим точно с една от хипотезите h_s и h_g. По този начин ако учителят каже, че примерът е положителен, тогава ще се промени S, а ако каже, че е отрицателен, ще се промени G. И в двата случая ще се намали пространството на версиите.
Това означава, че примерът $< 3, 3 >$ задължително ще намали пространството на версиите, а примерът $< 8, 8 >$ няма да го намали.

4. А сега да предположим, че сте учител, опитващ да научи алгоритъм на едно определено понятие (например $3 \leq x \leq 5, 2 \leq y \leq 9$). Какъв е най-малкият брой на обучаващите примери трябва да предоставите на алгоритъма за елиминиране на кандидати, за да може той абсолютно точно да научи това понятие?

   Решение За да се стигне до точно тази хипотеза, това означава, че G и S съвпадат и да се състоят точно от една хипотеза. Нека това е хипотезата:
   $$a \leq x \leq b, c \leq y \leq d$$
   Множеството S се ограничава като около положителните примери се построява най-малкият правоъгълник обграждащ примерите. За да се постигне правоъгълника отговарящ на търсената хипотеза само 2 примера са необходими: $< a, c >$ и $< b, d >$. След като те бъдат обработени S ще има вида:
   $$S = \{a \leq x \leq b, c \leq y \leq d\}$$
   Множеството G се ограничава като всеки един от отрицателните примери отсича една част от интервала за x или за y. Това означава, че за да се стигне от:
   $$-\infty \leq x \leq \infty, -\infty \leq y \leq \infty$$
   до
   $$a \leq x \leq b, c \leq y \leq d$$
   Трябва да получим четири отрицателни примера, които да ограничат x и y отгоре и отдолу. Това биха могли да бъдат
   $$\begin{array}{c} < a-1, t > \\ < b+1, t > \\ < t, c-1 > \\ < t, d+1 > \end{array}$$
   където t е произволно цяло число. По този начин G ще придобие вида:
   $$G = \{a \leq x \leq b, c \leq y \leq d\}$$