1 Problem 2.3

1.1 Initialization

I assume each attribute is binary-valued as assumed in our class notes.

$$G_0 = \{<?, ?, ?, ?, ?, ? > \vee <?, ?, ?, ?, ?, ? >\}$$

$$S_0 = \{< \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset > \vee < \emptyset, \emptyset, \emptyset, \emptyset, \emptyset, \emptyset >\}$$

1.2 Example 1

$$d =< s, w, n, s, w, s >$$

$$G_1 = G_0$$

$$S_1 = \\ < s, w, n, s, w, s > \vee < s, w, n, s, w, s >$$

1.3 Example 2

$$d =< s, w, h, s, w, s >$$

$$G_2 = G_1$$

$$S_2 = \\ < s, w, h, s, w, s > \vee < s, w, n, s, w, s >$$

1.4 Example 3

$$d = < r, c, h, s, w, c >$$

$$S_3 = S_2$$

Take inconsistent hypotheses in G and replace them with minimum generalizations of d. Fix G to

$$
\begin{array}{l}
G_3 = \\
\quad < s,?,?,?,?,? > \vee <?,w,?,?,?,? > \\
\quad < s,?,?,?,?,? > \vee <?,?,n,?,?,? > \\
\quad < s,?,?,?,?,? > \vee <?,?,?,w,?,? > \\
\quad < s,?,?,?,?,? > \vee <?,?,?,?,c,? > \\
\quad < s,?,?,?,?,? > \vee <?,?,?,?,?,s > \\
\quad <?,w,?,?,?,? > \vee <?,?,n,?,?,? > \\
\quad <?,w,?,?,?,? > \vee <?,?,?,w,?,? > \\
\quad <?,w,?,?,?,? > \vee <?,?,?,?,c,? > \\
\quad <?,w,?,?,?,? > \vee <?,?,?,?,?,s > \\
\quad <?,?,n,?,?,? > \vee <?,?,?,w,?,? > \\
\quad <?,?,n,?,?,? > \vee <?,?,?,?,c,? > \\
\quad <?,?,n,?,?,? > \vee <?,?,?,?,?,s > \\
\quad <?,?,?,w,?,? > \vee <?,?,?,?,c,? > \\
\quad <?,?,?,w,?,? > \vee <?,?,?,?,?,s > \\
\quad <?,?,?,?,c,? > \vee <?,?,?,?,?,s >
\end{array}
$$

such that some member of S is more specific than each member of G.

$$
\begin{array}{l}
G_3 = \\
\quad < s,?,?,?,?,? > \vee <?,w,?,?,?,? > \\
\quad < s,?,?,?,?,? > \vee <?,?,n,?,?,? >
\end{array}
$$

$$< s,?,?,?,?,? > \vee <?,?,?,w,?,? >$$
$$< s,?,?,?,?,? > \vee <?,?,?,?,c,? >$$
$$< s,?,?,?,?,? > \vee <?,?,?,?,?,s >$$
$$<?,w,?,?,?,? > \vee <?,?,n,?,?,? >$$
$$<?,w,?,?,?,? > \vee <?,?,?,w,?,? >$$
$$<?,w,?,?,?,? > \vee <?,?,?,?,c,? >$$
$$<?,w,?,?,?,? > \vee <?,?,?,?,?,s >$$
$$<?,?,n,?,?,? > \vee <?,?,?,?,?,s >$$
$$<?,?,?,w,?,? > \vee <?,?,?,?,?,s >$$
$$<?,?,?,?,c,? > \vee <?,?,?,?,?,s >$$

1.5 Example 4

$$d =< s,w,h,s,c,c >$$

Remove from G any hypotheses inconsistent with d.

$$G_4 =$$
$$< s,?,?,?,?,? > \vee <?,w,?,?,?,? >$$
$$< s,?,?,?,?,? > \vee <?,?,n,?,?,? >$$
$$< s,?,?,?,?,? > \vee <?,?,?,w,?,? >$$
$$< s,?,?,?,?,? > \vee <?,?,?,?,c,? >$$
$$< s,?,?,?,?,? > \vee <?,?,?,?,?,s >$$
$$<?,w,?,?,?,? > \vee <?,?,n,?,?,? >$$
$$<?,w,?,?,?,? > \vee <?,?,?,w,?,? >$$
$$<?,w,?,?,?,? > \vee <?,?,?,?,c,? >$$
$$<?,w,?,?,?,? > \vee <?,?,?,?,?,s >$$
$$<?,?,?,?,c,? > \vee <?,?,?,?,?,s >$$

$$S_4 =$$

$$< s, w, ?, s, w, s > \vee < s, w, n, s, c, c >$$

2 Problem 2.4

2.1 a,b.

The S-set is

$$4 <= x <= 6, 3 <= y <= 5.$$

The G-set is

$$2 <= x <= 7, 2 <= y <= 5; 3 <= x <= 7, 2 <= y <= 7.$$

Graphically, the set of most specific hypotheses (a unique set, in this case) is shown to the left of Figure 1, and the set of most general hypotheses is shown on the right.

2.2 c.

The point (4, 6) is guaranteed to reduce the size of the version space. It is strictly outside the S-set, and strictly inside the G-set. See Figure 2: if the trainer answers yes, then the S-set will expand to fit the diagram on the left; if the trainer answers no, the G-set will collapse to the one on the right.

On the other hand, any point that is either in the S-set (e.g., (5, 4) and (4, 3)) or outside the G-set (e.g. (8, 9)) will not narrow either one of the sets (assuming the data are consistent, and that the ẗrue" target concept can be represented in this hypothesis language).

2.3 d.

We can teach an S-set which corresponds exactly to the target concept by specifying any two diagonally opposite corners of the rectangle as the first two points, as shown in the figure.

To bound the G-set, we need to use four examples. Each example can cuf off at most half of an infinite plane, and we require four such cuts to make G exactly as specific as S, as shown in Figure 3. There are other ways to do this narrowing besides picking the negative points as shown, but all of them require four points.

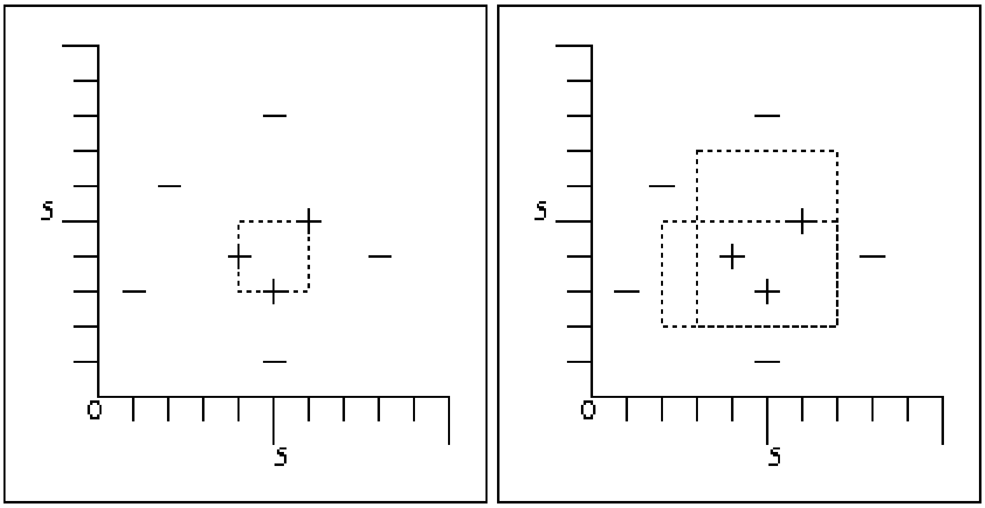

Figure 1: Most specific and most general hypotheses

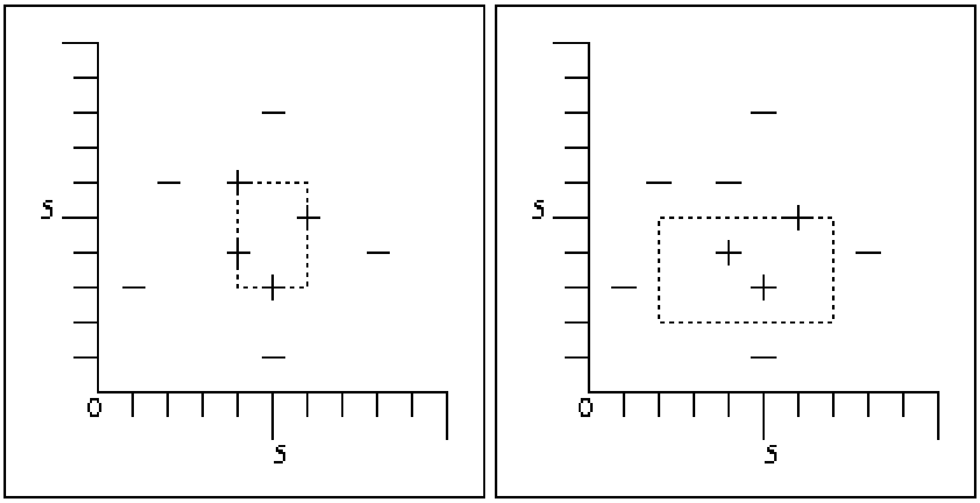

Figure 2: Result of testing (4,6) for + and - results