**Version Spaces** [Mitchell, 1977]

**The goal**
Produce a description consistent with all positive and no negative examples.

**The method**
Represent sets of descriptions by tracking:

\[ G = \text{the set of most general descriptions consistent with examples so far}. \]

\[ S = \text{the set of most specific descriptions consistent with examples so far}. \]

When the two meet (i.e., \( G = S \)), you know what the concept is.

**Candidate Elimination Algorithm**

1. Initialize \( G \) to the universal description (all variables) and \( S \) to contain the first positive example.

2. **Process next example:**
   - **If positive,**
     - (a) Discard inconsistent elements of \( G \).
     - (b) Replace inconsistent elements of \( S \) with their minimal generalizations in the version space that include the example.
     - (c) Discard any element of \( S \) more general than another element of \( S \).
   - **If negative,**
     - (a) Remove elements of \( S \) that cover the example.
     - (b) Replace elements of \( G \) that cover the example with their minimal specializations in the version space that exclude the example.
     - (c) Discard any element of \( G \) less general than another element of \( G \).
   - 3. If \( G \) and \( S \) are singletons and equal, halt. If \( G \) and \( S \) are singletons and not equal, give up.
      Else go to 2.

Example sequence:

+ (Japan, Honda, blue, 1980, economy)
- (Japan, Toyota, green, 1970, sports)
+ (Japan, Toyota, blue, 1990, economy)
- (USA, Chrysler, red, 1980, economy)
+ (Japan, Honda, white, 1980, economy)

For this example, we assume

- The representation language cannot represent disjunctions.
- There are no abstractions of the given features except for variables.

First example is positive:

(Japan, Honda, blue, 1980, economy)

Initialize \( G \) to \{ \( (x_1, x_2, x_3, x_4, x_5) \) \}

Initialize \( S \) to

\{ (Japan, Honda, blue, 1980, economy) \}
Second example is negative:
(Japan, Toyota, green, 1970, sports)

We specialize $G$ to exclude the example, but remain in the version space:

$G = \{ (x_1, Honda, x_3, x_4, x_5),$
\hspace{1cm} $(x_1, x_2, blue, x_4, x_5),$
\hspace{1cm} $(x_1, x_2, x_3, 1980, x_5),$
\hspace{1cm} $(x_1, x_2, x_3, x_4, economy) \}$

$S = \{ (Japan, Honda, blue, 1980, economy) \}$

(it’s unchanged by this negative example)

---

Third example is positive:
(Japan, Toyota, blue, 1990, economy)
Previous, we had

$G = \{ (x_1, Honda, x_3, x_4, x_5),$
\hspace{1cm} $(x_1, x_2, blue, x_4, x_5),$
\hspace{1cm} $(x_1, x_2, x_3, 1980, x_5),$
\hspace{1cm} $(x_1, x_2, x_3, x_4, economy) \}$

We remove from $G$ the descriptions that are inconsistent with this positive example, getting:

$G = \{ (x_1, x_2, blue, x_4, x_5),$
\hspace{1cm} $(x_1, x_2, x_3, x_4, economy) \}$

---

Fourth example is negative:
(USA, Chrysler, red, 1980, economy)
We previously had

$G = \{ (x_1, x_2, blue, x_4, x_5),$
\hspace{1cm} $(x_1, x_2, x_3, x_4, economy) \}$

We specialize the elements of $G$ that are consistent with this example, to rule the example out but still permit prior positive examples.

$G = \{ (Japan, x_2, blue, x_4, x_5),$
\hspace{1cm} $(Japan, x_2, x_3, x_4, economy) \}$

$S$ is unaffected.
Fifth example is positive:
(Japan, Honda, white, 1980, economy)

Previously G was

\[ G = \{ (x_1, x_2, \text{blue}, x_4, x_5), \\ (\text{Japan}, x_2, x_3, x_4, \text{economy}) \} , \]

We remove from G all the descriptions that don’t cover the example, getting

\[ G = \{ (\text{Japan}, x_2, x_3, x_4, \text{Economy}) \} \]

Previously we had
\[ S = \{ (\text{Japan}, x_2, \text{Blue}, x_4, \text{Economy}) \} \]
which doesn’t include
(Japan, Honda, white, 1980, economy)

The minimal generalization to include it is:
\[ S = \{ (\text{Japan}, x_2, x_3, x_4, \text{economy}) \} \]
Since \( G = S \) are singletons, we’re done.
The concept learned: ”Japanese economy car.”