Disadvantage :-

→ consider only 'tve' values

→ it is not guarantee that it will match all the data because we ignored '-ve' samples. (most general hypothesis)

(Find - S algorithm)

Candidate elimination algorithm :-

To overcome the disadvantage of Find-S algorithm, candidate elimination algorithm is used. (Both positive and negative instance can be considered)

Consistent hypothesis :-

An hypothesis is said to be consistent hypothesis with a set of training example Dataset (D) if $h(x) = C(x)$

$h(x)$ → hypothesis of instance

$C(x)$ → concept of instance

Version space :-

→ It is an intermediate of General hypothesis & Specific hypothesis

→ It gives all the possible hypothesis.

Candidate elimination algorithm :-

→ Accepts both positive & negative instance

→ Uses "Version Space"

+ve instance ⇒ Generalize Specific hypothesis

-ve instance ⇒ Make General hypothesis more specific

Algorithm :-

→ Initilize 'G' general hypothesis & 'S' specific hypothesis.

→ For each instance

if instance is +ve

Generalize specific hypothesis

∴ (Find-S)

else

Make general hypothesis more specific

"Considering the same dataset used for Find-S algorithm"

Example :- Same dataset used for Find-S algorithm.

Step - 1 :-

Initilize 'G' and 'S'

$G = \{'?', '?', '?', '?', '?', '?'\}$

$S = \{'\phi', '\phi', '\phi', '\phi', '\phi', '\phi'\}$

Step - 2 :-

First instance → Positive instance

$S = \{$'sunny', 'warm', 'Normal', 'strong', 'warm', 'same'$\}$

G → will remain same,

$G = \{'?', '?', '?', '?', '?', '?'\}$

Second instance → Positive instance

'S' becomes,

$S = \{$'sunny', 'warm', '?', 'strong', 'warm', 'same'$\}$

$G = \{'?', '?', '?', '?', '?', '?'\}$

Third instance ⟶ negative instance

→ 'S' is retained / remain same

$G_1 = \{ ?, ?, ?, ?, ?, ? \}$

~~For each attribute in 'G1' to all~~

For each attribute in 'G_1',

$G_1 \Rightarrow \{$ <sunny, ?, ?, ?, ?, ?> ✓

<?, Warm, ?, ?, ?, ?> ✓

<?, ?, Normal, ?, ?, ?> ✗

<?, ?, ?, Cool, ?> ✗

<?, ?, ?, ?, '?', Same> ✓ $\}$

Wind → attribute remains same so it doesn't consider.

If all the attribute in G_1 are consistent with all instance we can consider in 'G_1'.

$G_1 = \{$ <sunny, ?, ?, ?, ?, ?>,

<?, Warm, ?, ?, ?, ?>,

<?, ?, ?, ?, ?, Same> $\}$

Fourth instance → Positive instance

S = {'sunny', 'warm', '?', 'strong', '?', 'same'}

~~The resultant 'S' contradicting with~~

G ⇒ remains same

For finding all the possible hypothesis we use Version space

S: ① <sunny, warm, ?, strong, ?, same>

② <sunny, warm, ?, ?, ?, ?> <sunny, ?, ?, strong, ?, ?>

③ <sunny, ?, ?, ?, ?, same> <?, warm, ?, strong, ?, ?>

④ <?, warm, ?, ?, ?, same> <sunny, ?, ?, ?, ?, same>

⑤ <?, ?, ?, strong, ?, same>

⑥

G: <sunny, ?, ?, ?, ?, ?> <?, warm, ?, ?, ?, ?>

<?, ?, ?, ?, ?, 'same'>

9 ⇒ Consistent hypothesis.