GENERAL BOUNDARY, SPECIFIC BOUNDARY VERSION SPACE REPRESENTATION THEOREM:

Definition: The general boundary G, with respect to hypothesis space H and training data D, is the set of maximally general members of H consistent with D.

G = { g ∈ H | consistent (g, D) ^ (-∃g' ∈ H)[(g' >₉ g )^ consistent (g', D)]}

Definition: The specific boundary S, with respect to hypothesis space H and training data D, is the set of minimally general (i.e., maximally specific) members of H consistent with D.

G = { g ∈ H | consistent (s, D) ^ (-∃s' ∈ H)[(s >₉ s' )^ consistent (s', D)]}

Version space representation theorem. Let X be an arbitrary set of instances and let H be a set of Boolean-valued hypotheses defined over X.

Let c : X → {O, 1} be an arbitrary target concept defined over X, and let D be an arbitrary set of training examples {<x, c(x)>}. For all X, H, c, and D such that S and G are well defined,

VSʜ,ᴅ = { h ∈ H |(∃s ∈ S)(g ∈ G)(g ≥₉ h ≥₉ s )}

Proof: To prove the theorem it suffices to show that (1) every h satisfying the right hand side of the above expression is in VSʜ,ᴅ and (2) every member of VSʜ,ᴅ satisfies the right-hand side of the expression. To show (1) let g be an arbitrary member of G, s be an arbitrary member of S, and h be an arbitrary member of H, such that g ≥₉ h ≥₉ s. Then by the definition of S, s must be satisfied by all positive examples in D. Because

h ≥₉ s, h must also be satisfied by all positive examples in D. Similarly, by the definition of G, g cannot be satisfied by any negative example in D, and because g ≥₉ h, h cannot be satisfied by any negative example in D. Because h is satisfied by all positive examples in D and by no negative examples in D, h is consistent with D, and therefore h is a member of VSʜ,ᴅ .

This proves step (1). The argument for (2) is a bit more complex. It can be proven by assuming some h in VSʜ,ᴅ that does not satisfy the right-hand side of the expression, then showing that this leads to an inconsistency.