Jericho McLeod
G00986513

Ex 2.1:

The hypothesis space in EnjoySport is 973 because there are six attributes with a total of 18 possible values in the form: [3, 2, 2, 2, 2, 2]

A hypothesis space also contains "?" values meaning any possible value for the attribute is acceptable for a hypothesis, and a "Ø" value indicating no possible value for the attribute is acceptable.

Representing this in the vector showing possible values, the Ø only needs to appear once, as any appearance is semantically identical. This means that semantically distinct combinations are limited to:
[4,3,3,3,3,3]+1 possible combinations, as the "?" value is added to the possible values and the "Ø" is added a single time.

(4*3*3*3*3*3)+1 = 973.

An additional attribute, *A*, with *k* possible values, impacts the hypothesis space, *H*, as:
$H_1 = ((H_0-1)*(k+1))+1$
Or, the new hypothesis space, H_1, is equal to the old Hypothesis space, H_0, less the one possible expression containing any number of "Ø" values (-1), times the new variable's quantity of possible values plus the possibility that any value will satisfy the hypothesis (*(k+1)), then adding back the possibility of a single "Ø" in any position, (+1).

In the example in the book, this is reflected by taking:
973-1 = 972
972 * 4 = 3,888
3,888+1 = 3,889, which is the new hypothesis space for an additional attribute, WaterCurrent, with possible values [Light, Moderate, Strong].

Ex 2.2:

S0 = [Ø,Ø,Ø,Ø,Ø,Ø]
G0 = [?,?,?,?,?,?,?]
S1 = [Sunny, Warm, High, Strong, Cool, Change]
G1 = [?,?,?,?,?,?,?]
S2 = [Sunny, Warm, ?, ?, Cool, ?]
G2 = [Sunny,?,?,?,?,?,?], [Cloud,?,?,?,?,?,?] ,[?,Warm,?,?,?,?,?] ,[?,?,?,?,?,Cool,?]
S3 = [Sunny, Warm, ?, ?, ?, ?]
G3 = [Sunny,?,?,?,?,?,?], [Cloud,?,?,?,?,?,?] ,[?,Warm,?,?,?,?,?]
S4 = [Sunny, Warm, ?, ?, ?, ?]
G4 = [Sunny,?,?,?,?,?,?], [Cloud,?,?,?,?,?,?] ,[?,Warm,?,?,?,?,?]

The final boundary sets will be identical because they are sets determined from each instance. If a set is [1,1] and you subtract one from a value, and then another, the final set [0,0] exists regardless of which value is changed first.

In this set, placing example 4 before example 3 would mean the general set would not split into a collection of 4 sets, which would reduce the size of the boundary sets if the training occurred in the original order.

Ex 2.5 (a,b):
Part A:

Data Dictionary for Table:

- Sex:
  - M = Male
  - F = Female
- Hair:
  - Bl = Black
  - Br = Brown
  - Blo = Blonde
- Height:
  - T = Tall
  - M = Medium
  - S = Short
- Nation:
  - US = US
  - FR = French
  - DE = German
  - IN = Indian
  - IE = Ireland
  - JP = Japanese
  - PT = Portugese

Person	P1				P2				Result
Attr	Sex	Hair	Height	Nation	Sex	Hair	Height	Nation	
Obs1	M	Br	T	US	F	Bl	S	US	+
Obs2	M	Br	S	FR	F	Bl	S	US	+
Obs3	F	Br	T	DE	F	Bl	S	IN	-
Obs4	M	Br	T	IE	F	Br	S	IE	+

S0 = [<Ø,Ø,Ø,Ø>,<Ø,Ø,Ø,Ø>]
G0 = [<?,?,?,?>,<?,?,?,?>]

S1 = [<M,Br,T,US>,<F,Bl,S,US>]
G1 = [<?,?,?,?>,<?,?,?,?>]

S2 = [<M,Br,?,?>,<F,Bl,S,US>]
G2 = [<?,?,?,?>,<?,?,?,?>]

S3 = [<M,Br,?,?>,<F,Bl,S,US>]
G3 = [<M,?,?,?>,<?,?,?,?>],

[<?,?,?,US>,<?,?,?,?>], [<?,?,?,FR>,<?,?,?,?>], [<?,?,?,IN>,<?,?,?,?>],
[<?,?,?,IE>,<?,?,?,?>], [<?,?,?,JP>,<?,?,?,?>], [<?,?,?,PT>,<?,?,?,?>]

[<?,?,?,?>,<?,?,?,US>], [<?,?,?,?>,<?,?,?,FR>], [<?,?,?,?>,<?,?,?,DE>],
[<?,?,?,?>,<?,?,?,IE>], [<?,?,?,?>,<?,?,?,JP>], [<?,?,?,?>,<?,?,?,PT>]

S4 = [<M,BR,?,?>,<F,?,S,?>]
G4 = [<M,?,?,?>,<?,?,?,?>],

[<?,?,?,US>,<?,?,?,?>], [<?,?,?,FR>,<?,?,?,?>], [<?,?,?,IN>,<?,?,?,?>],
[<?,?,?,IE>,<?,?,?,?>], [<?,?,?,JP>,<?,?,?,?>], [<?,?,?,PT>,<?,?,?,?>]

[<?,?,?,?>,<?,?,?,US>], [<?,?,?,?>,<?,?,?,FR>], [<?,?,?,?>,<?,?,?,DE>],
[<?,?,?,?>,<?,?,?,IE>], [<?,?,?,?>,<?,?,?,JP>], [<?,?,?,?>,<?,?,?,PT>]

Part B:

Given the set [<M, Bl, S, PT>, <F, Blo, T, IN>], there can be no “Ø” values. A hypothesis containing any other value than those listed would not be consistent, thus the only acceptable values are those shown and “?”.

2^8 = 256

Ex 2.7:

There are infinitely many numbers between 4.5 and 6.1.

If I were to claim that 4.51 is the lowest number > 4.5, I could disprove my claim by showing that 4.501 is > 4.5, then show an even lower number 4.5001 is > 4.5.

The use of <= or => instead of inequalities alone would solve this.