2 Concept Learning

Consider the hypothesis space defined over these instances, in which each hypothesis is represented by a pair of 6-tuples, and where each attribute constraint may be a specific value, "?", or "Ø". The task is to learn whether the guy will go to play or not. Please provide a hand trace of the Candidate-Elimination algorithm learning from the example in Table 1 and hypothesis language. In particular, show the specific and general boundaries of the version space after it has processed the first training example, then second example, etc.

Example	Sky	Temperature	Water	Humidity	Day	Forecast	Play
1	Sunny	Hot	Light	High	Weekend	Same	Yes
2	Sunny	Mild	Strong	High	Weekday	Change	Yes
3	Sunny	Mild	Moderate	Low	Weekday	Change	Yes
4	Rainy	Hot	Strong	Low	Weekend	Change	No
5	Rainy	Mild	Moderate	Low	Weekday	Same	No
6	Sunny	Mild	Light	High	Weekend	Same	Yes

Table 1: Positive and Negative training examples for the target concept PlayTennis

5. Candidate Elimination Algorithm (CEA)

Consider a concept learning problem where the data D, which concerns ancient Egyptian vases discovered in archeological excavations, is expressed as tuples of five attributes:

damaged, color, material, kingdom, markings.

Examples are classified as either valuable (+) or not valuable (-), and D consists of the following:

Numb	er	Example					
_ 1	<no,< th=""><th>brown,</th><th>marble,</th><th>middle,</th><th>hieroglyphics></th><th><u>+</u></th></no,<>	brown,	marble,	middle,	hieroglyphics>	<u>+</u>	
_ 2	<no,< td=""><td>white,</td><td>sandstone,</td><td>old,</td><td>none></td><td></td></no,<>	white,	sandstone,	old,	none>		
_ 3	<no,< td=""><td>white,</td><td>marble,</td><td>new,</td><td>hieroglyphics></td><td>+</td></no,<>	white,	marble,	new,	hieroglyphics>	+	
4	<yes< td=""><td>, grey,</td><td>slate,</td><td>middle,</td><td>hieroglyphics></td><td>-</td></yes<>	, grey,	slate,	middle,	hieroglyphics>	-	
_ 5	<no,< td=""><td>brown,</td><td>granite,</td><td>middle,</td><td>hieroglyphics></td><td>+</td></no,<>	brown,	granite,	middle,	hieroglyphics>	+	

Assume that all possible values of each attribute are represented in D above.

- (a) What is the size of the hypothesis space searched by the candidate elimination algorithm (CEA) using the data D given above?
- (b) Suppose the CEA has seen examples 1 and 2 only so far. Show its current specific boundary S_2 and general boundary G_2 for the version space.
- (c) Show S₃ and G₃ after the CEA also sees example 3.
- (d) Show S_5 and G_5 after the CEA also sees the final two examples 4 and 5.