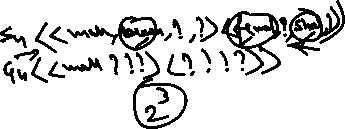
**Assignment 1. Exercise 2.5 of Tom Mitchell Book:**

Consider the following sequence of positive and negative training examples describing the concept "pairs of people who live in the same house." Each training example describes an ***ordered*** pair of people, with each person described by their sex, hair, color (black, brown, or blonde), height (tall, medium, or short), and nationality (US,French, German, Irish, Indian, Japanese, or Portuguese).

+ ((male brown tall US) (female black short US))

+ ((male brown short French)( female black short US))



- ((female brown tall German)( female black short Indian))



+ ((male brown tall Irish) (female brown short Irish))

Consider a hypothesis space defined over these instances, in which each hypothesis is represented by a pair of 4-tuples, and where each attribute constraint may be a specific value, "?," or "0," just **as** in the EnjoySport hypothesis representation. For example, the hypothesis

((male ? tall ?)(female ? ? Japanese))

represents the set of all pairs of people where the first is a tall male (of any nationality and hair color), and the second is a Japanese female (of any hair color and height).

1. Provide a hand trace of the CANDIDATE-ELIMINATION algorithm learning from the above training examples and hypothesis language. In particular, show the specific and general boundaries of the version space after it has processed the first training example, then the second training example, etc.
2. How many distinct hypotheses from the given hypothesis space are consistent with the following single positive training example?

+ ((male black short Portuguese)(female blonde tall Indian))



1. Assume the learner has encountered only the positive example from part ***(b),*** and that it is now allowed to query the trainer by generating any instance and asking the trainer to classify it. Give a specific sequence of queries that assures the learner will converge to the single correct hypothesis, whatever it may be (assuming that the target concept is describable within the given hypothesis language). Give the shortest sequence of queries you can find. How does the length of this sequence relate to your answer to question (b)?
2. Note that this hypothesis language cannot express all concepts that can be defined over the instances (i.e., we can define sets of positive and negative examples for which there is no corresponding describable hypothesis). If we were to enrich the language so that it could express all concepts that can be defined over the instance language, then how would your answer to (c) change?