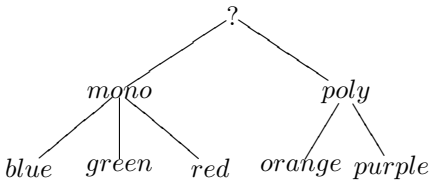
Session 3: Version Spaces

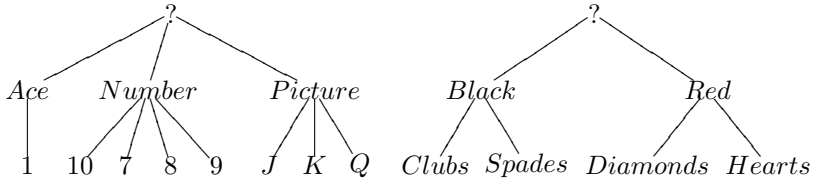
1. *Version Spaces: Colors.* Consider the following concept hierarchy from the course notes:



Apply the Version-Space algorithm with the following examples:

red: + purple: - blue: +

2. *Version Spaces: Playing cards.* We consider playing cards to have two characteristics: "value" (7, 8, 9, 10, Jack (J), Queen (Q), King (K), Ace) and "kind" (Hearts, Diamonds, Spades, Clubs). As such, we have the following two concept hierarchies:



Several concepts (models) can be described using these two hierarchies. E.g. [?,?] is the most general concept, [?,red] is more specific and [number,red] and [?,Diamonds] are even more specific. The concept [9,Diamonds] is maximally specific (individual card) and cannot be further specialized. Consider the following examples:

7 of Diamonds: + Ace of Clubs: - Queen of Hearts: -
9 of Hearts: + 8 of Clubs: -

Apply the Version-Space algorithm.

3. *Version spaces: Ex-exam.* A rock band is searching for a new drummer. They place an advertisment in "Rolling Stone" magazine, and get 900 applications from drummers. All candidates hand in a CV, including a description of their hobbies, favourite music, age, and a picture. The band doesn't want to interview 900 persons, so they select 5 candidates from the 900, and select the good drummers out of these 5. Then, they use the information from these 5 examples to decide (using the version spaces algorithm) which drummers of the 900 candidates are the most interesting, such that they only have to interview a part of the candidates.

   The examples are:

Hobby	music	age	photo	selected ?
fishing	hiphop	16	handsome	no
stage-diving	newwave	18	neutral	yes
dancing	hardrock	32	ugly	no
music-only	hardrock	25	handsome	yes
stage-diving	jazz	29	ugly	no

   The hobbies, ages and favourite music are organised in general-to-specific hierarchies, given in the figures below. This means that e.g. the hypothesis [boring. young, ?, ?] is more general than [boring, 19,not-useful,?], which in its turn is more general than [fishing, 19, not-useful, handsome]. Apply the Version Space algorithm on this example. Indicate per example how the spaces G and S evolve. Clearly indicate which hypotheses are pruned away, and why.

   What is (or what are) the concept(s) that are deducted ?

   Using this result, what conclusion can the band make for the following additional CVs (should the band interview the drummer or not, or is it undecided ?)

   [music-only, hardrock, 32, handsome]
   [stage-diving, hiphop, 18, neutral]
   [dancing, new-wave, 22, ugly]

4. We consider a very primitive computer screen. It is square in shape and has 36 pixels. We can identify every pixel by its horizontal and vertical coordinates: for all natural numbers x and y, with $0 \leq x, y < 6$, there is a location (x, y). In every location, three light sources can be activated: a red one, a blue one, and a green one. The screen only allows us to show squares. The sides of squares are parallel to the X- and Y-axis of the screen. Their corners are locations (x, y) of the screen; in particular, the length of the sides of a square is a natural number. We can only show complete squares (that is, we cannot show squares which are partly outside of the screen). A *coloured square* is represented by a pair $[((x, y), n), colour]$.

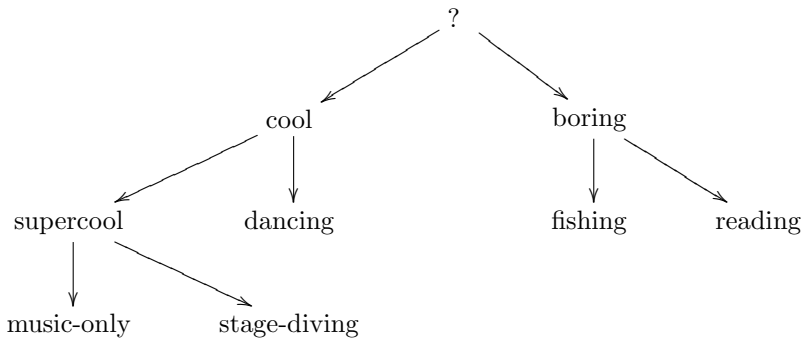


Figure 1: Hobby hierarchy for Example 3.

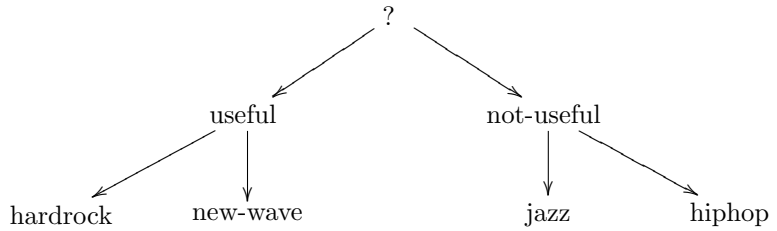


Figure 2: Music hierarchy for Example 3.

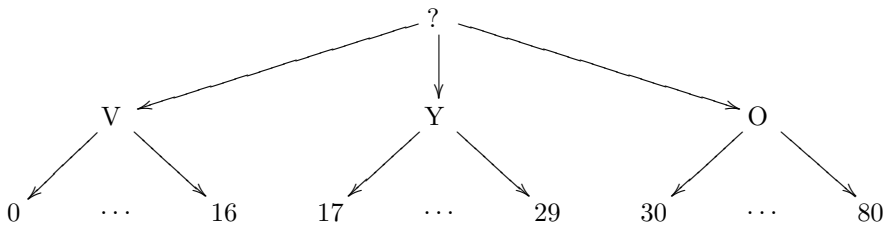


Figure 3: Age hierarchy (V=very-young, Y=young, O=old) for Example 3.

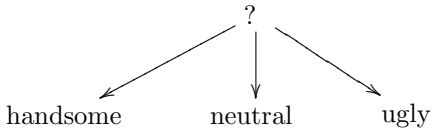


Figure 4: Picture hierarchy for Example 3.

Here, $0 \leq x, y < 6$ are the X- and Y-coordinates of the bottom-left corner of the square and n is the length of its sides. So, the corners of the square are the locations: $(x, y), (x, y + n), (x + n, y), (x + n, y + n)$. The squares are closed, that is, they include the border. The second property of a coloured square, its *colour*, is one of the values in the colour concept hierarchy shown in Figure 5. In this hierarchy, for instance, violet is more general than red, because purple is shown using both the red and the blue light source. **Remark: a point is not a square! So,** $n > 0$**.**

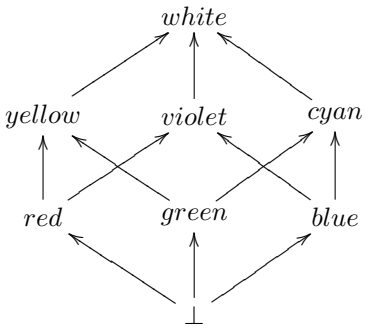


Figure 5: Colour hierarchy for Example 4

The way in which we make a square appear on the screen is by giving a number of examples of light sources that should, or shouldn't, be lit. We consider the following examples:

Location	Colour	Lit?
(0,1)	red	no
(3,2)	red	yes
(1,4)	green	yes
(4,5)	red	no
(4,0)	green	no

Use the Version Spaces algorithm to determine which coloured square could be shown on the screen after these five examples. **The hypothesis language is that of coloured squares, extended with the** $\perp$**-symbol.** Show step-by-step how the sets G and S evolve. Point out what is pruned and why. The examples must be dealt with in the order in which they are given above.

Can you determine whether the following light sources will be lit?

$$[(3, 2), green] \qquad [(2, 4), red] \qquad [(3, 5), blue]$$