Multimedia II: Machine Learning & Computer Vision SS18

Assignment 02

Multimedia Computing Lab
Prof. Dr. Rainer Lienhart
M. Sc. Dan Zecha

Submitting your solution is not mandatory. Your solutions will be discussed in the exercise on May 10th, 2019.

Exercise 1 (Hypotheses and Hypothesis Spaces)

1. Explain why the hypothesis space H in the EnjoySport learning task contains 973 semantically distinct hypotheses. How many syntactically distinct hypotheses are within H?

2. How would the number of possible instances and possible hypotheses increase with the addition of the attribute *WaterCurrent*, which can take on the values *None*, *Light*, *Moderate*, or *Strong*?

3. More generally, how does the number of possible instances and hypotheses grow with addition of a new attribute A that takes on k possible values?

Exercise 2 (CANDIDATE-ELIMINATION)

Let a class of geometric objects be described by the attributes

- **color** with the possible values *red*, *green*, and *blue*

- **shape** with the possible values *cube*, *sphere*, and *pyramid*

- **size** with the possible values *small*, *medium*, and *large*

- **weight** with the possible values *light* and *heavy*

1. What is the size of the hypothesis space?

2. Give a trace of the CANDIDATE-ELIMINATION algorithm for the following sequence of training examples:

Instance	Color	Shape	Size	Weight	Classification
1	Red	Cube	Small	Heavy	−
2	Blue	Pyramid	Small	Light	+
3	Green	Sphere	Medium	Light	−
4	Red	Pyramid	Large	Light	+

3. Explain why the final version space output by the CANDIDATE-ELIMINATION algorithm will always be the same regardless of the sequence of a given set of training examples!

4. The sets S and G computed at intermediate stages will, of course, depend on this sequence. Can you come up with ideas for ordering the training examples to minimize the sum of the sizes of these intermediate S and G sets for the H used in the geometric objects example?