

Assignment: Decision Trees

Problem 1

Questions:

- What is a decision tree?
- How is a decision tree applied to an input x ?
- How is a decision tree learned?
- What is the importance / usefulness of an attribute?
- What is feature engineering in NLP?

Problem 2

Give examples for the following concepts.

- Important / useful attribute (say for which task)
- Useless attribute (say for which task)

Problem 3

Your task is to engineer features for a decision tree that takes a name as input and returns either the label “Arab” if the name is a traditional Arab name or the label “German” if the name is a traditional German name.

Describe one feature (in English or German, not as a program) that is likely to be useful for this decision tree. Why is it useful?

Problem 4

Look at the following training set for a decision tree that classifies names as either location or person names:

name	attribute 1 first name	attribute 2 geo-term	label
George Washington Bridge	yes	yes	location
John Hancock Tower	yes	yes	location
Ronald Reagan Building	yes	yes	location
Azore Islands	no	yes	location
Michael Jackson	yes	no	person
Justin Bieber	yes	no	person
Lady Gaga	no	no	person
Adele	yes	no	person
Stalin	no	no	person

Which attribute is more useful, attribute 1 or attribute 2? Why?

Problem 5

Below you see a decision tree for location vs. person. Indicate how this tree processes (1) "AZORE ISLANDS", (2) "LONDON", (3) "MAIER" and (4) "JOHN MAIER" by drawing a line from the root to a leaf node. Mark each line with its number (1, 2, 3 or 4) at the leaf node. Which names are correctly classified, which incorrectly?

