

Homework Assignment 5

Exercise 5.1

What is the F measure used for? What is its underlying intuition? Do not use any formulas.

Exercise 5.2

Where does the sawtooth shape come from when drawing precision–recall curves for ranked lists?

Exercise 5.3

What does the Cluster Hypothesis say? How can it be exploited for information retrieval tasks?

Exercise 5.4

How can one determine whether a clustering is “good”?

Exercise 5.5

What is the idea underlying the k-means clustering algorithm? Using your own words, briefly explain how the algorithm works. Do not use pseudocode.

Exercise 5.6

What is a dendrogram and how can it be used in clustering tasks?