Optimal decision tree pipeline: proof of concept.

Phases:

1) Maximal hyperbox clustering

2) Minimal cutset

3) Minimal decision tree

So far five testsets: four tiny artificial (few points, 2 dimensions), fifth is iris dataset (https://archive.ics.uci.edu/dataset/53/iris), attributes only Sepal Width and Petal Length, well known to separate iris setosa from the other two species (trivial).

Phase 1 (optional), AABB clustering

Identifies AABB (hyperbox) clusters of points belonging to one same class. Reasonable when points are not too intermixed (which would make clustering meaningless, and the whole analysis of little relevance).

Directory hyperbox, two implementations: C++ (complete), python (incomplete, but produces plots).

C++: testset file must be written inside the code, file hyperbox.cpp.

Results will be in file hyperboxes\_TESTSETNAME.txt

Phase 2: minimal cutset identification

Identifies the minimal set of cuts that separates exactly all dataset points.

Directory MIPmodel, only python implementation.

Test1

A graph with numbers and dots

Description automatically generated

A diagram with red lines and dots

Description automatically generated

A graph with colored lines and dots

Description automatically generated

Test2

A graph with numbers and dots

Description automatically generated

A red rectangle with black dots and yellow dots

Description automatically generated

A graph with numbers and dots

Description automatically generated

Test 3

A graph with numbers and dots

Description automatically generated

A red squares with yellow and purple dots

Description automatically generated

A graph with colored lines and dots

Description automatically generated

Test 4

A graph with numbers and dots

Description automatically generated

A screenshot of a graph

Description automatically generated

A graph with colored lines and dots

Description automatically generated

Iris, dimensions: SepalWidthCm, PetalLengthCm

A graph with numbers and dots

Description automatically generated

A graph with numbers and dots

Description automatically generated

A graph showing numbers and points

Description automatically generated