Ordered Sets, Big Homework

Task:

To build a classifier over FCA method and to compare it with the common algorithms.

Data Set: banknote authentication <https://archive.ics.uci.edu/ml/datasets/banknote+authentication>

Number of instances: 1372

Train: 75%, Test: 25%

4 continuous attributes

1 binary target

Using histogram, I took thresholds to transform continuous attributes to binary.

First algorithm is lazy FCA.

For each test instance I calculate votes for class "+" and for class "-".

if 𝑣𝑜𝑡𝑒𝑠"+"==0 and 𝑣𝑜𝑡𝑒𝑠"-"==0, there is no prediction for such instance.

if 𝑣𝑜𝑡𝑒𝑠"+"/(1+𝑣𝑜𝑡𝑒𝑠"−")<𝑡𝑎𝑢 this instance belongs to class +, otherwise to class -.

𝑡𝑎𝑢 – is a hyperparameter.

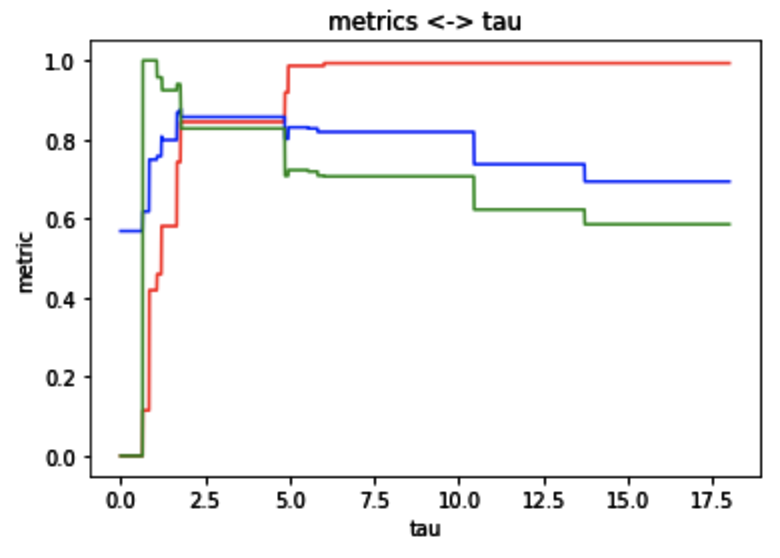
The second realized algorithm is the lazy FCA with online approach. Each test instance appends to the train set after evaluating.

Metrics for first algo:

Red – recall

Blue – accuracy

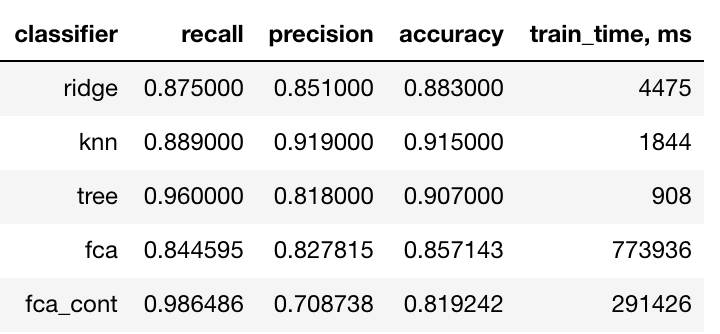
Green – precision



Best tau – is 3.

Metrics for Lazy FCA with online train is almost the same. For it took best tau = 6

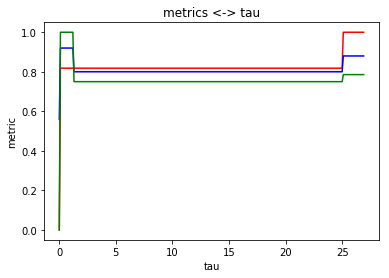
The result metrices table:



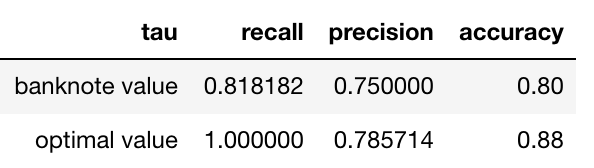
Let’s use it for Iris dataset.

Only 1-st and 2-nd classes.

Graph shows dependency metrics of tau



Best tau for banknote’s task shows worse result, than the optimal one (tau=26).



Results: 1) For my datasets FCA couldn’t win any of classic ml algorithms.

2) Best hyperparameter for one task isn’t the best one for another one.