Lazy FCA Classification

Chekmareva Kristina,
Vodopyanova Yulia

Dataset

- **Features**:

<u>Categorical</u>	<u>Numerical</u>

  - workclass
  - education
  - marital-status
  - occupation
  - relationship
  - race
  - gender
  - native-country

  - age
  - educational-num
  - capital-gain
  - capital-loss
  - hours-per-week

- **Target** — 'income':

  >50K → **1**   <=50K → **0**

- **Shape**:

  11208 rows; 14 columns

Classification improvement №1

1   Pandas Loops   →   NumPy (Vectorized computation)

2   Bool as pd.Series   →   Bit Packing (using uint64 blocks)

3   Counting classifiers   →   Weighing classifiers

4   —   →   Minimum support threshold

5   —   →   Top-K strongest classifiers

6   —   →   Built-in explainability

Classification improvement №2

1 Accounting for all training samples → Only NearestNeighbors

2 No normalization → Standartization (StandardScaler)

3 Counting classifiers → Weighing classifiers: $(\exp(-\alpha * dist))$

4 Global support → Local support (across neighbors)

Improved vs Initial versions

	f1-score (weighted)	f1-score	accuracy	time
FCA_NN_2 (improved)	0.8596	0.8667	0.860	309.356047
FCA_1 (improved)	0.8495	0.8585	0.850	15.397997
FCA (vanilla)	0.8341	0.8465	0.835	448.871114

Overall metrics comparison

Model	Accuracy	F1	ROC-AUC	TP	TN	FP	FN
LogReg	0.875	0.876847	0.93950	89	86	14	11
Lazy FCA Improved (NN)	0.860	0.866667	0.86000	91	81	19	9
Lazy FCA Improved	0.850	0.858491	0.85000	91	79	21	9
Naive Bayes	0.845	0.854460	0.94110	91	78	22	9
Random Forest	0.850	0.854369	0.92260	88	82	18	12
Decision Tree	0.845	0.853081	0.89805	90	79	21	10
SVM RBF	0.845	0.851675	0.92030	89	80	20	11
Lazy FCA	0.835	0.846512	0.83500	91	76	24	9
XGBoost	0.830	0.834951	0.92430	86	80	20	14
KNN	0.820	0.828571	0.89990	87	77	23	13

Interpretability

Built-in explanatory function explain_sample():

ex.:

Positive classifiers: 39
Negative classifiers: 187

=== Classifier information: ===
Strength: 2.0986
Support: +0 / -2
Binary features:
  - workclass_Private: True
  - education_Some-college: True
  - marital-status_Never-married: True
  - race_White: True
  - gender_Male: True
Numerical intervals:
  - age: [24.00, 24.00]
  - educational-num: [10.00, 10.00]
  - capital-gain: [0.00, 0.00]
  - capital-loss: [0.00, 0.00]
  - hours-per-week: [40.00, 40.00]

Thank you :)