RULE COVERING FOR INTERPRETATION & BOOSTING

Ilker Birbil

ERASMUS UNIVERSITY ROTTERDAM

OUTLINE

INTERPRETATION OF RANDOM FORESTS

PREPARATION FOR ASSIGNMENT 1

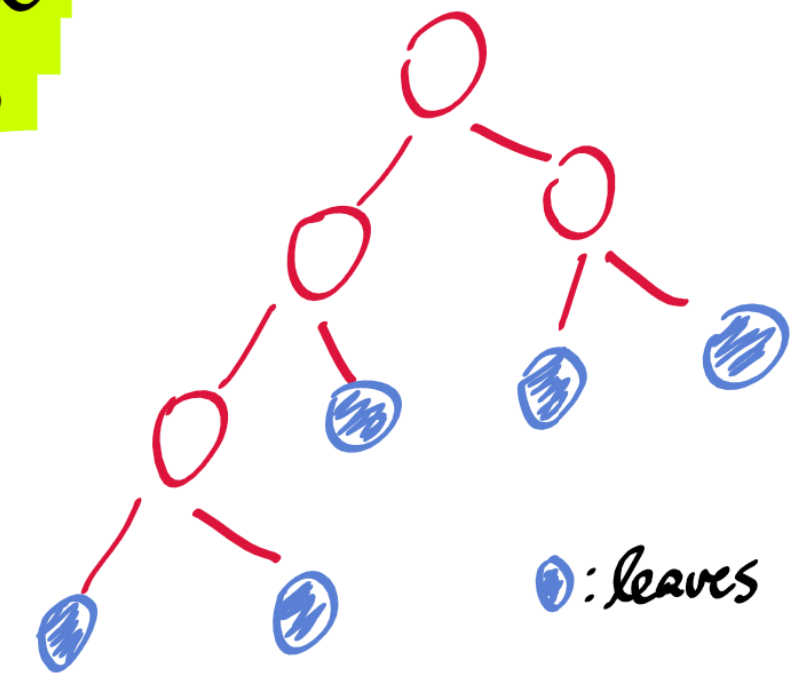
Data: $\{\triangle, \triangle, \bigcirc, \square, \triangle, \bigcirc, \bigcirc, \square\}$

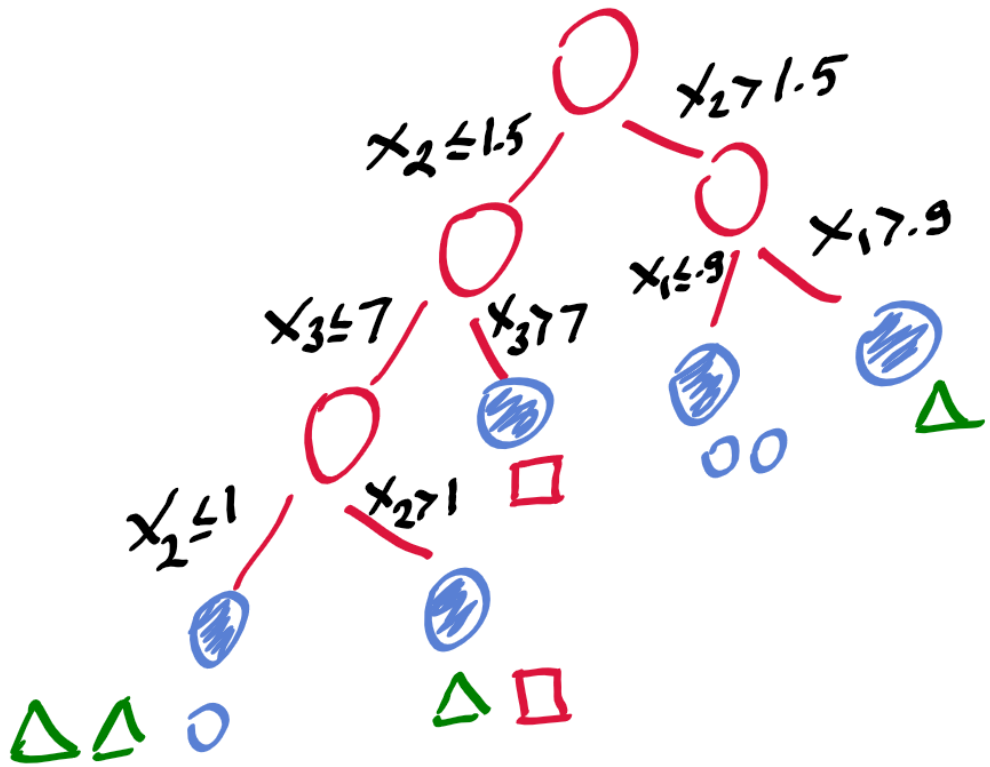
continuous features

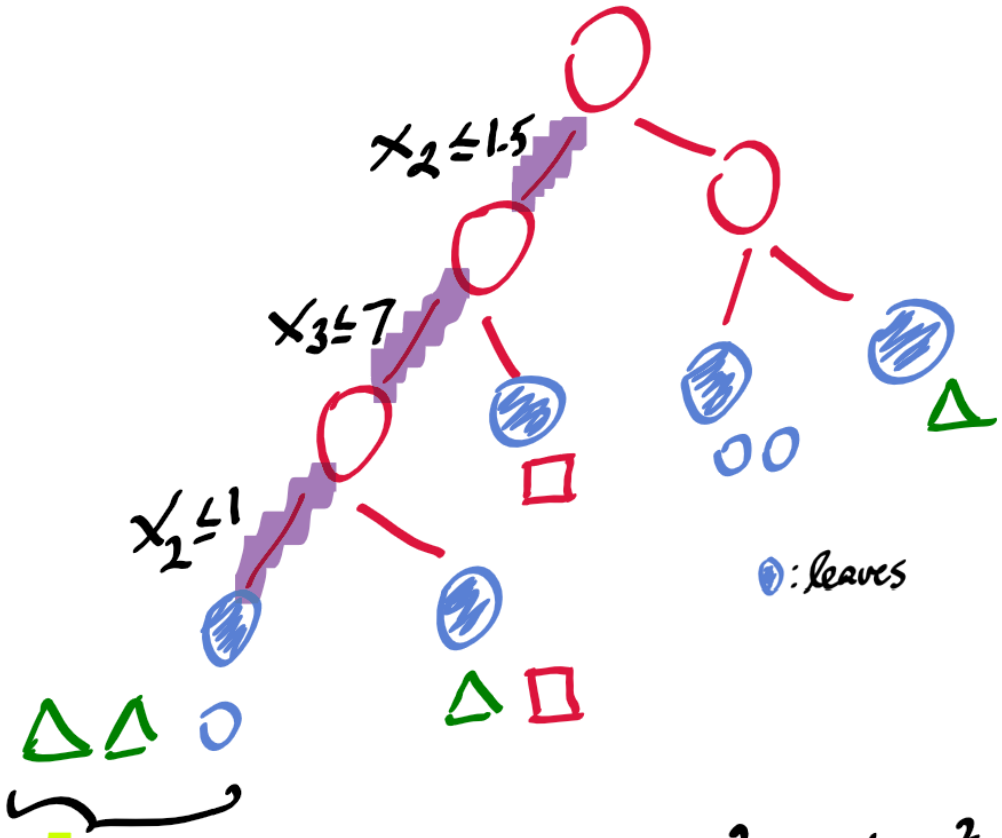
$x_1, x_2, \ldots, x_8$

$\square$ ← $\triangle$ ← $\bigcirc$ ← labels

DECISION TREES







Rule j (leaf): $x_2 \le 1$ & $x_3 \le 7$

Out of sample  ?

① Check the rules (leaves)

② Classify with **majority voting**

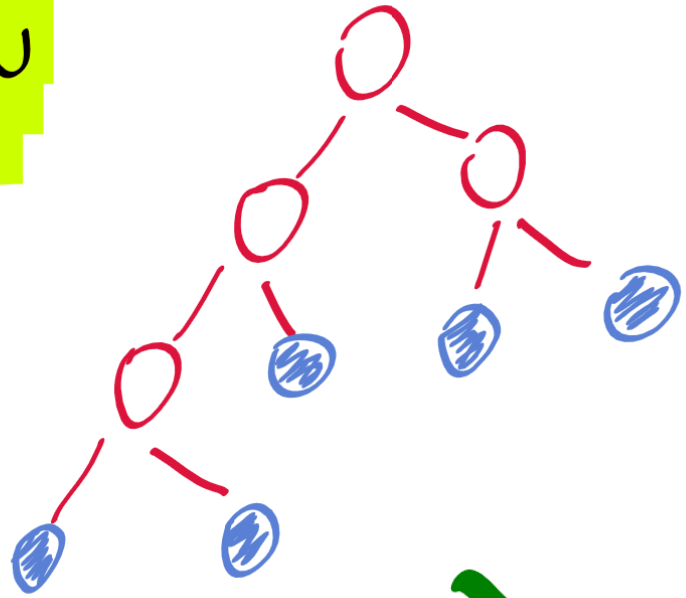
Impurity: $1 - \left(\frac{2}{3}\right)^2 - \left(\frac{1}{3}\right)^2$ — Gini Criterion

w_j

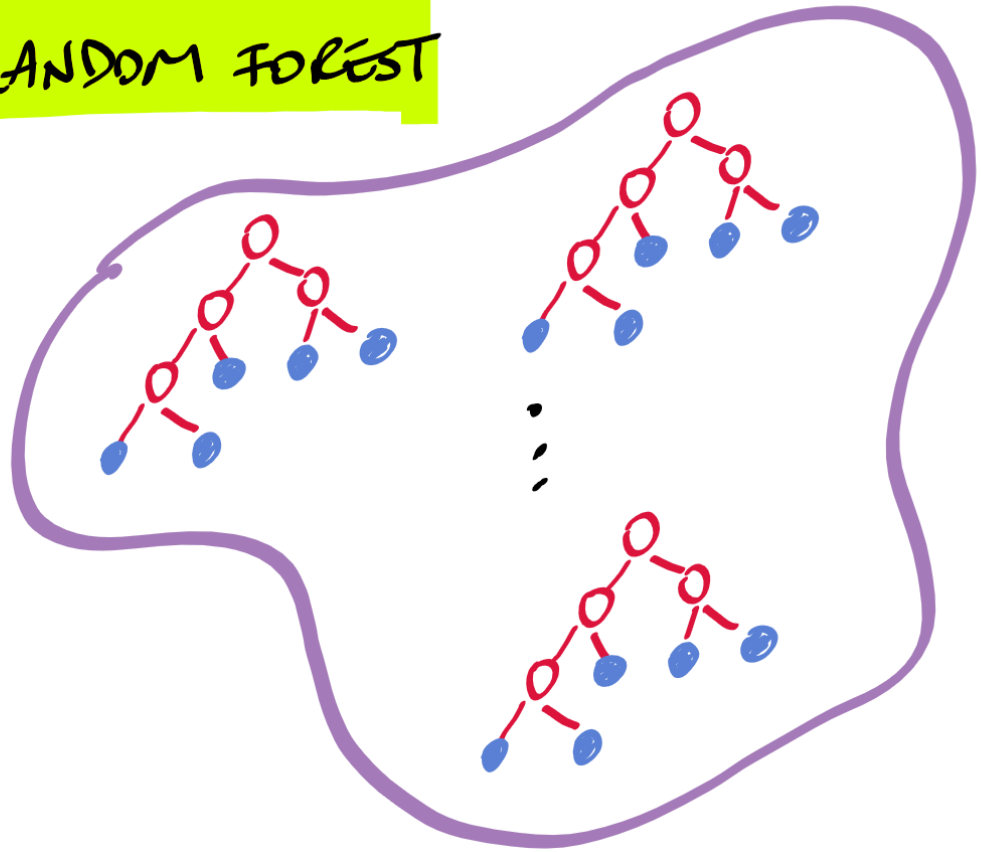
used for **splitting**

Ilker Birbil, 2020

DECISION TREES

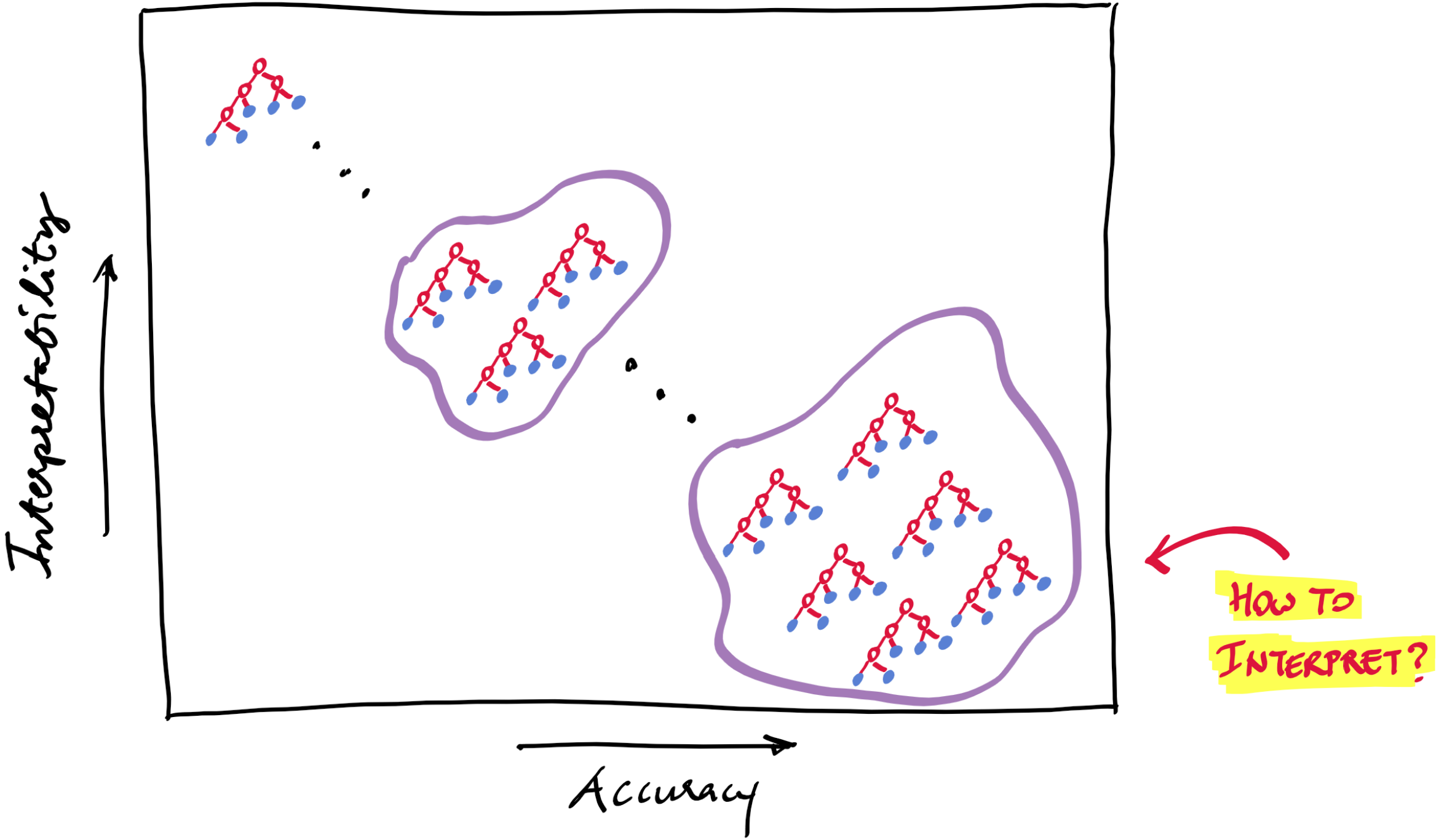


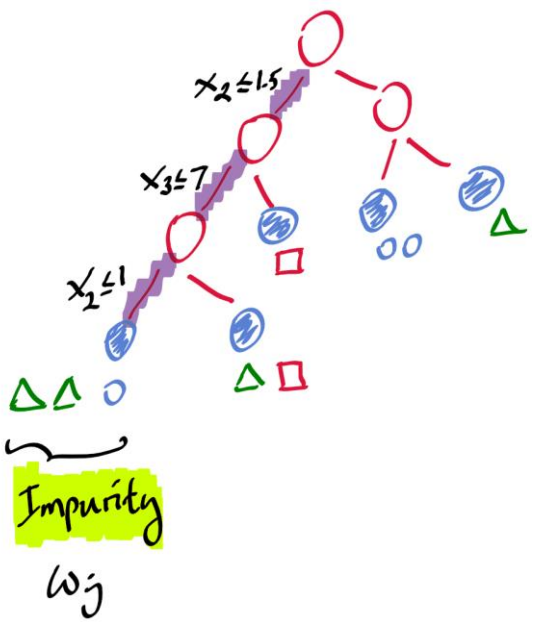
RANDOM FOREST



VARIANCE ↓ ACCURACY ↑

INTERPRETABILITY ↓





Data: $\{\triangle, \triangle, \circ, \square, \triangle, \circ, \circ, \square\}$

$$\begin{matrix} \triangle \\ \triangle \\ \circ \\ \square \\ \triangle \\ \circ \\ \circ \\ \square \end{matrix} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{bmatrix}$$

← Subset covered with rule j

J: set of all rules in a trained random forest

Decision to select rule j: $z_j \in \{0,1\}$

Cost of rule j: w_j

$$\text{Minimize} \quad \sum_{j \in J} (1 + w_j) z_j$$

$$\text{subject to} \quad \sum_{j \in J(i)} z_j \geq 1, \quad i \in I$$

$$z_j \in \{0,1\}$$

Rules covering sample i

SET COVERING PROBLEM

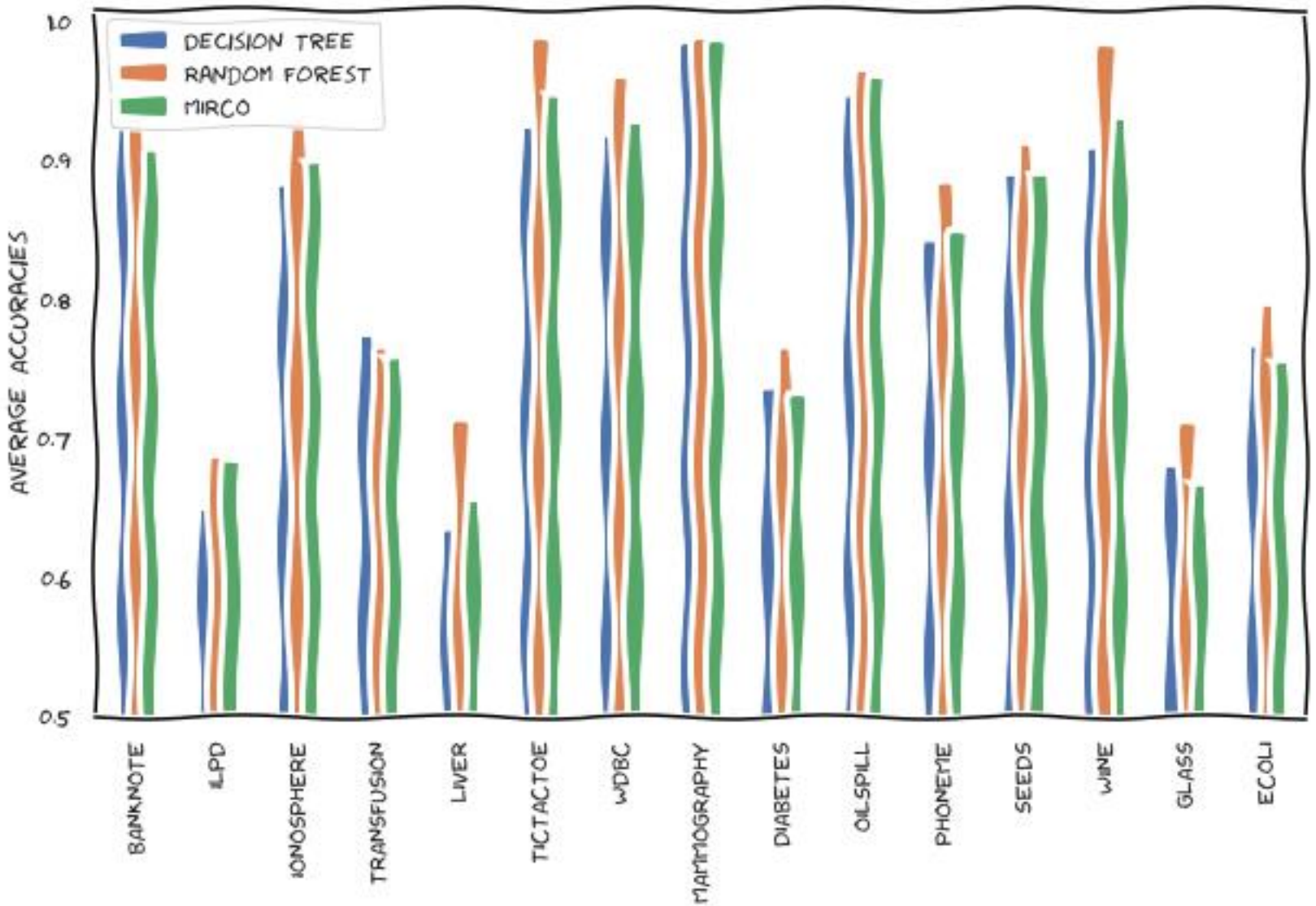
NP-HARD

Solvers ☹

Heuristics ☺

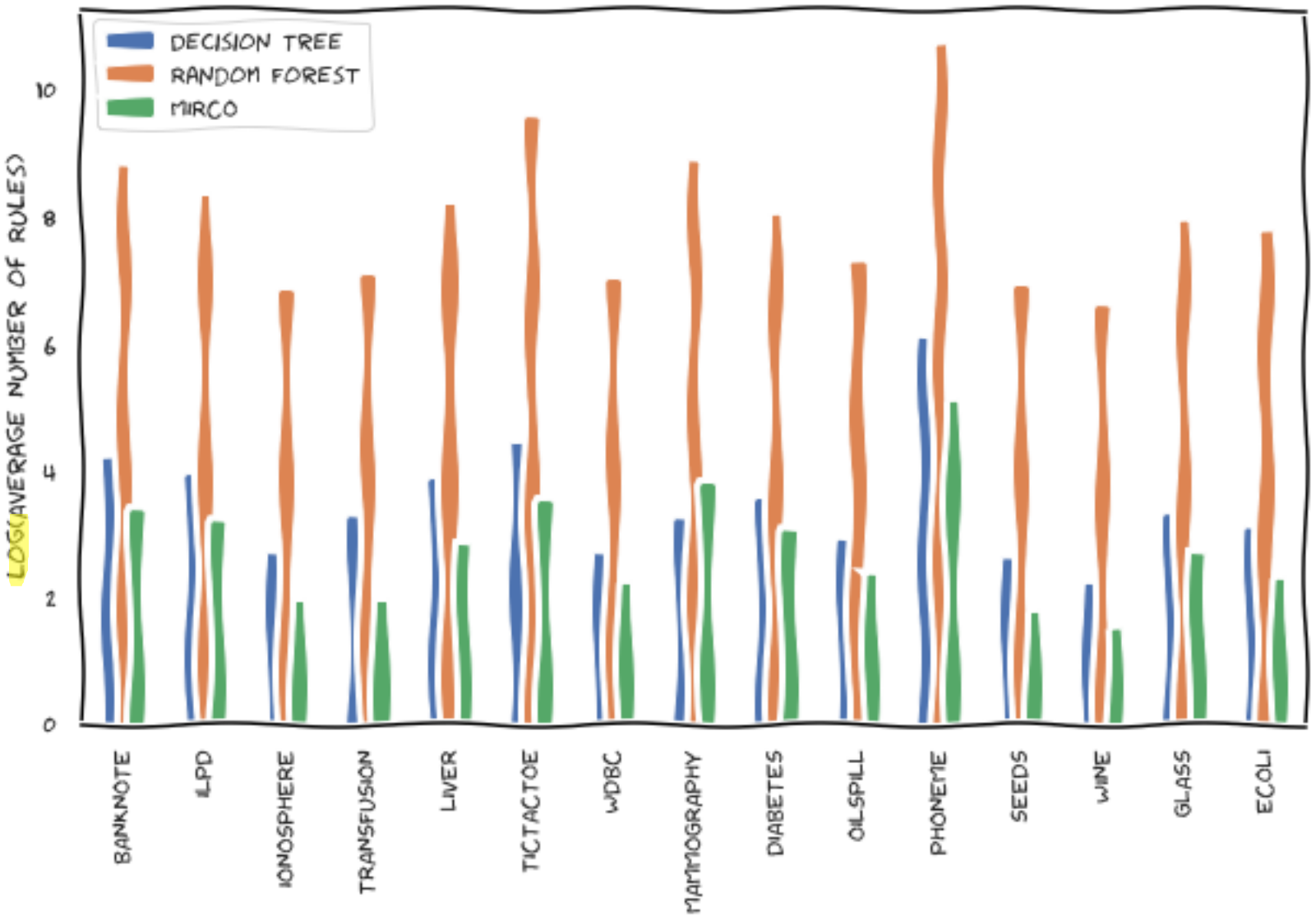
MINIMUM RULE COVER (MIRCO)

COMPUTATIONAL STUDY



Read the paper about missed test samples...

COMPUTATIONAL STUDY



IMPLEMENTATION

```
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import train_test_split
from MIRCO import MIRCO
```

⋮

```
RF = RandomForestClassifier(max_depth=5)
RF_fit = RF.fit(X, y)
RF_pred = RF_fit.predict(X)
```

⋮

```
MRC = MIRCO(RF_fit)
MRC_fit = MRC.fit(X, y)
MRC_pred = MRC_fit.predict(X)
```

```
356                 retdict['colLeafNos'][col] = leafno
357                 retdict['colTreeNos'][col] = treeno
358                 col += 1
359
360         return retdict
```

Only one file

CONCLUSION

INTERPRETATION OF RANDOM FORESTS WITH RULE COVERING $\Rightarrow$ MIRCO

ASSIGNMENT 1