KU LEUVEN DTAI Declarative Languages & Artificial Intelligence

A Ranking Stability Measure for Quantifying the Robustness of Anomaly Detection Methods

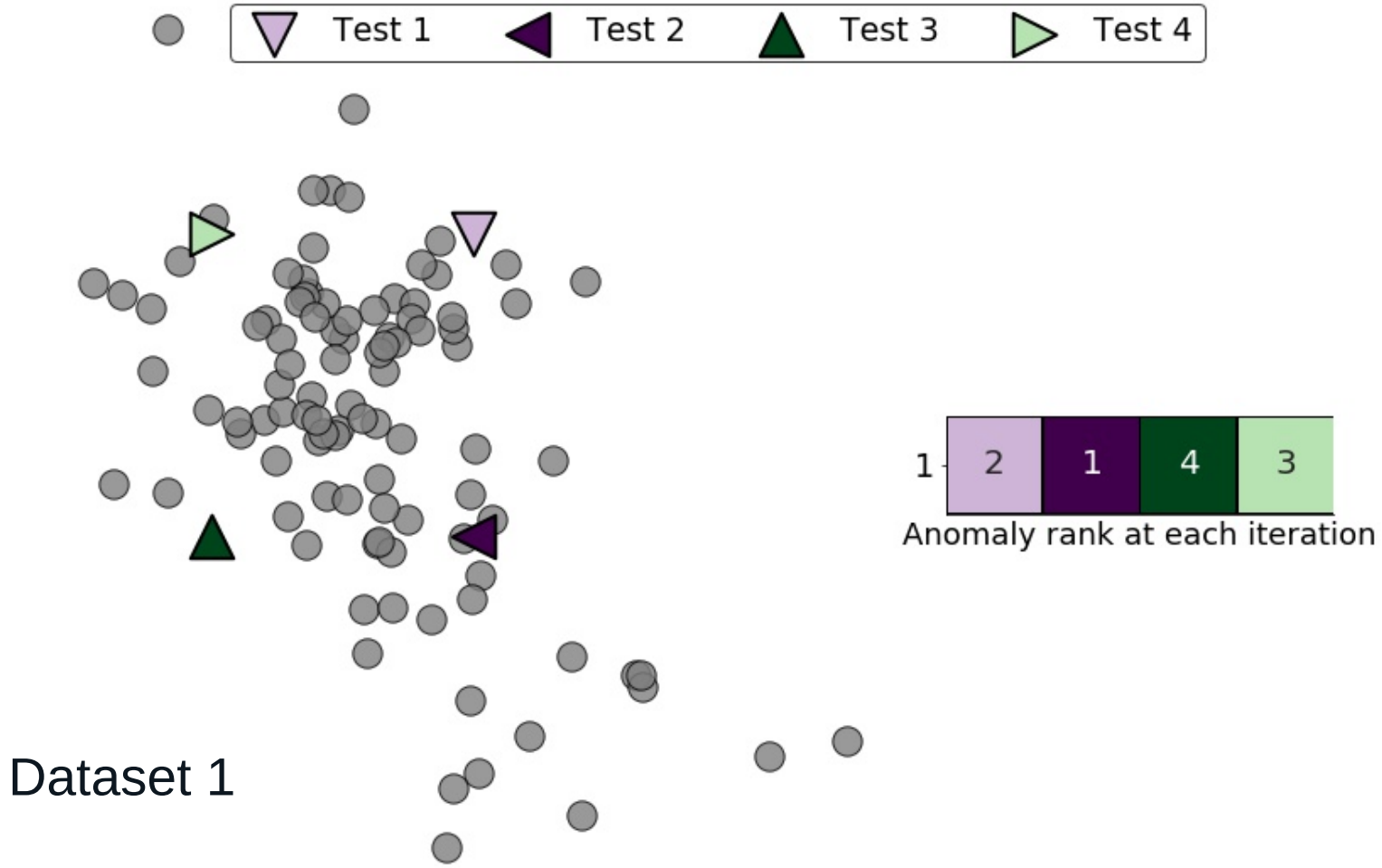
Lorenzo Perini, Connor Galvin, Vincent Vercruyssen

EDML 2020

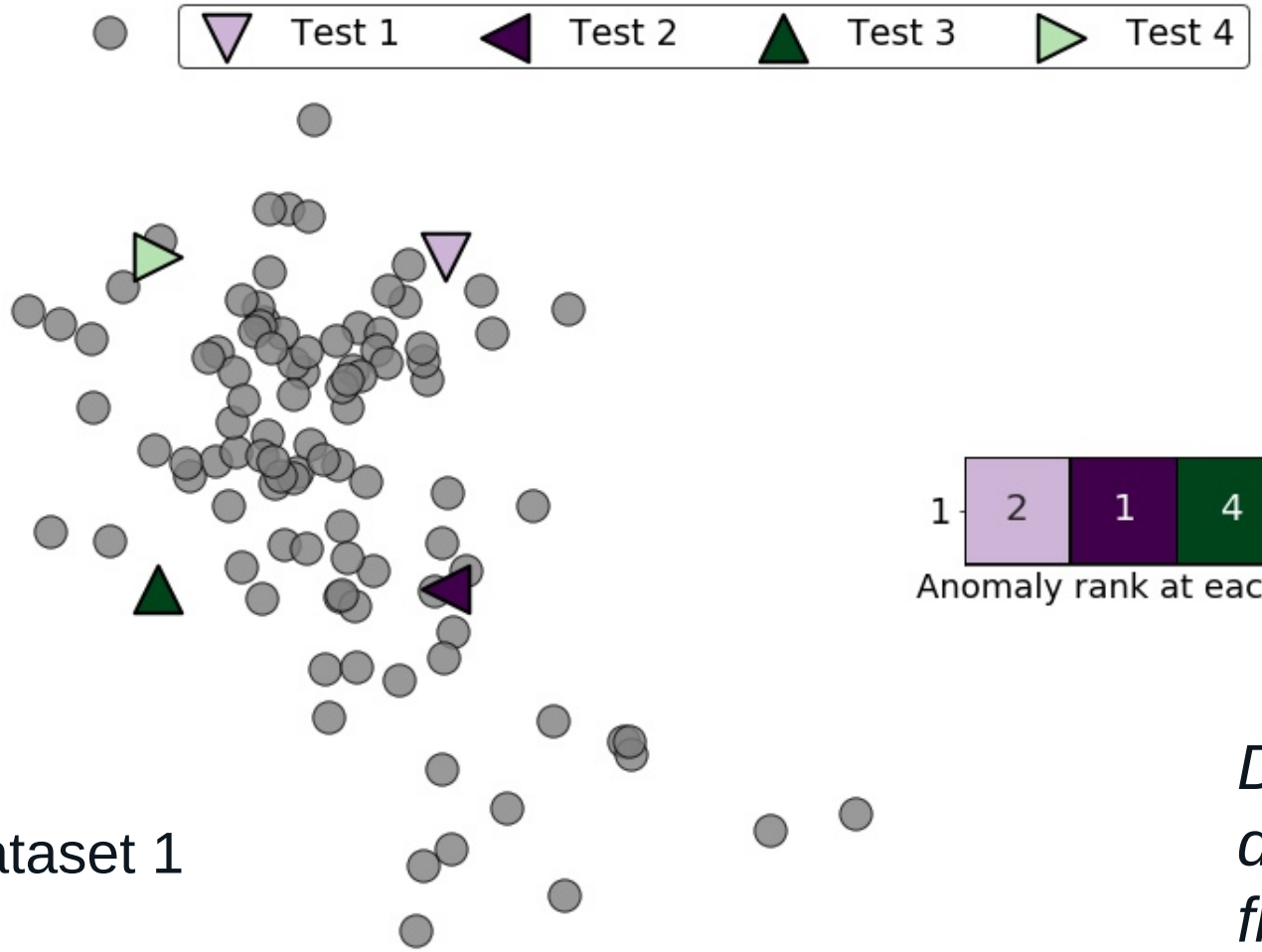
https://people.cs.kuleuven.be/~lorenzo.perini/

@LorenzoPerini95

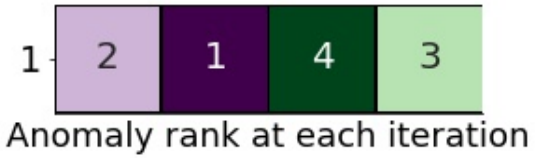
Capturing Variations in Anomaly Rankings



Capturing Variations in Anomaly Rankings

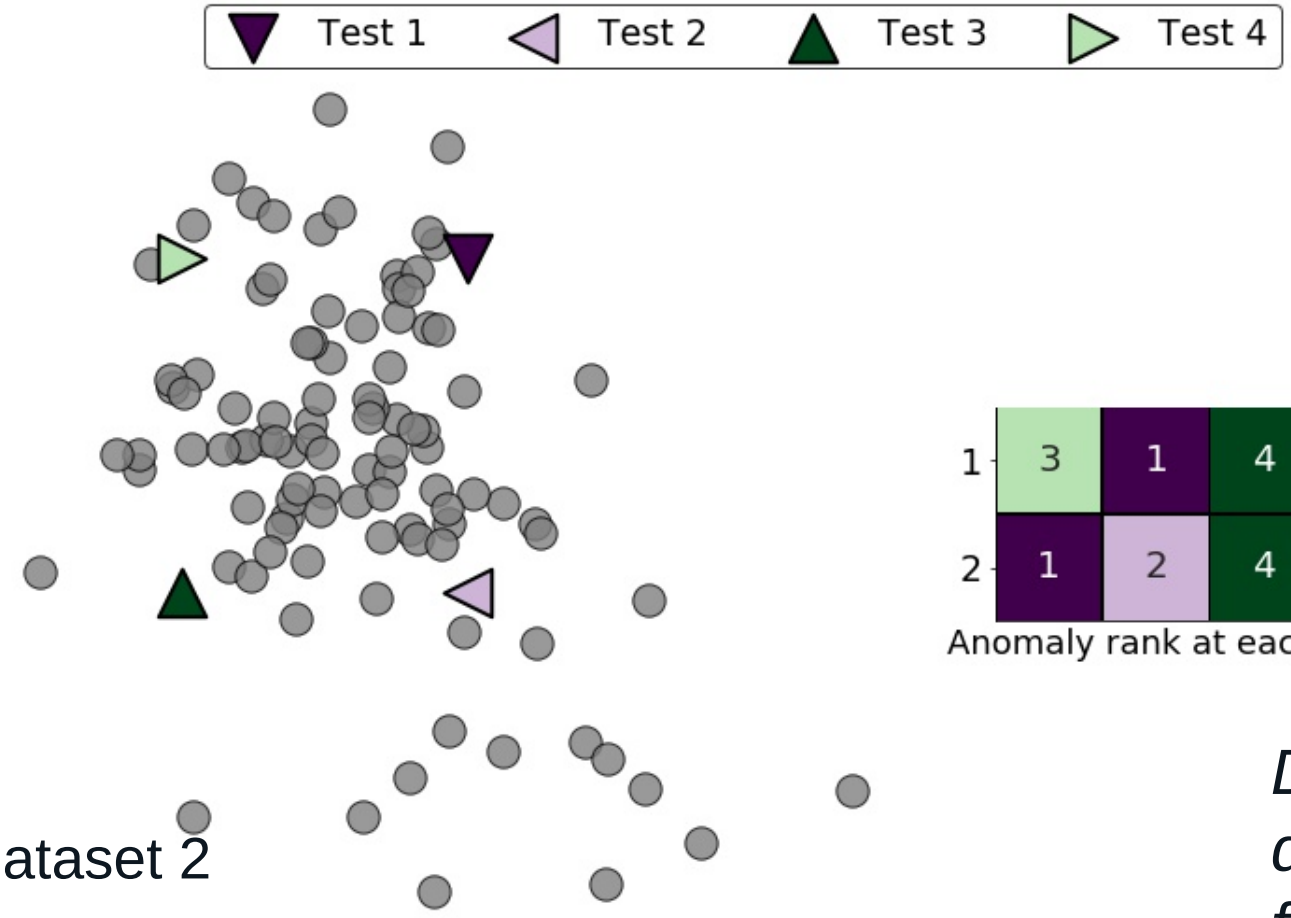


Dataset 1



Does the ranking change when drawing different training sets from the same distribution?

Capturing Variations in Anomaly Rankings

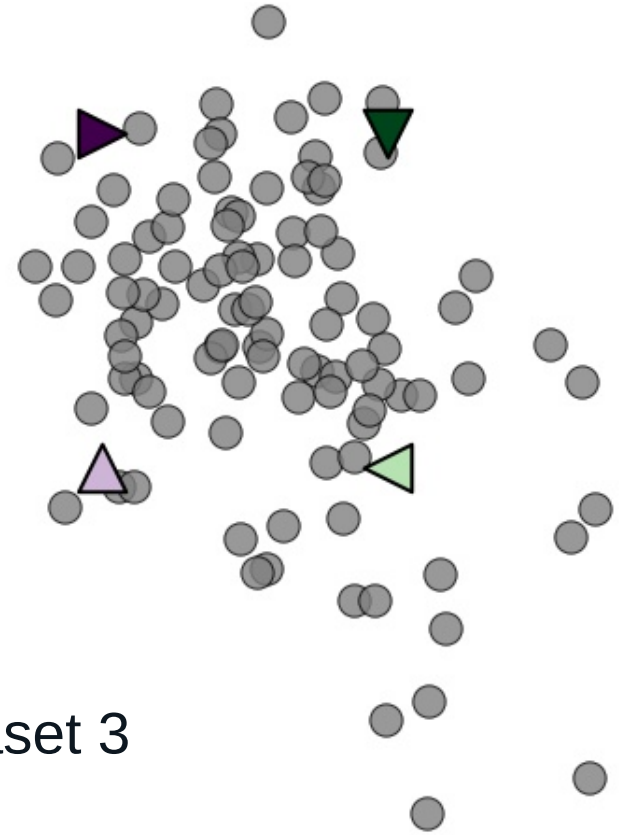


Dataset 2

Anomaly rank at each iteration

Does the ranking change when drawing different training sets from the same distribution?

Capturing Variations in Anomaly Rankings



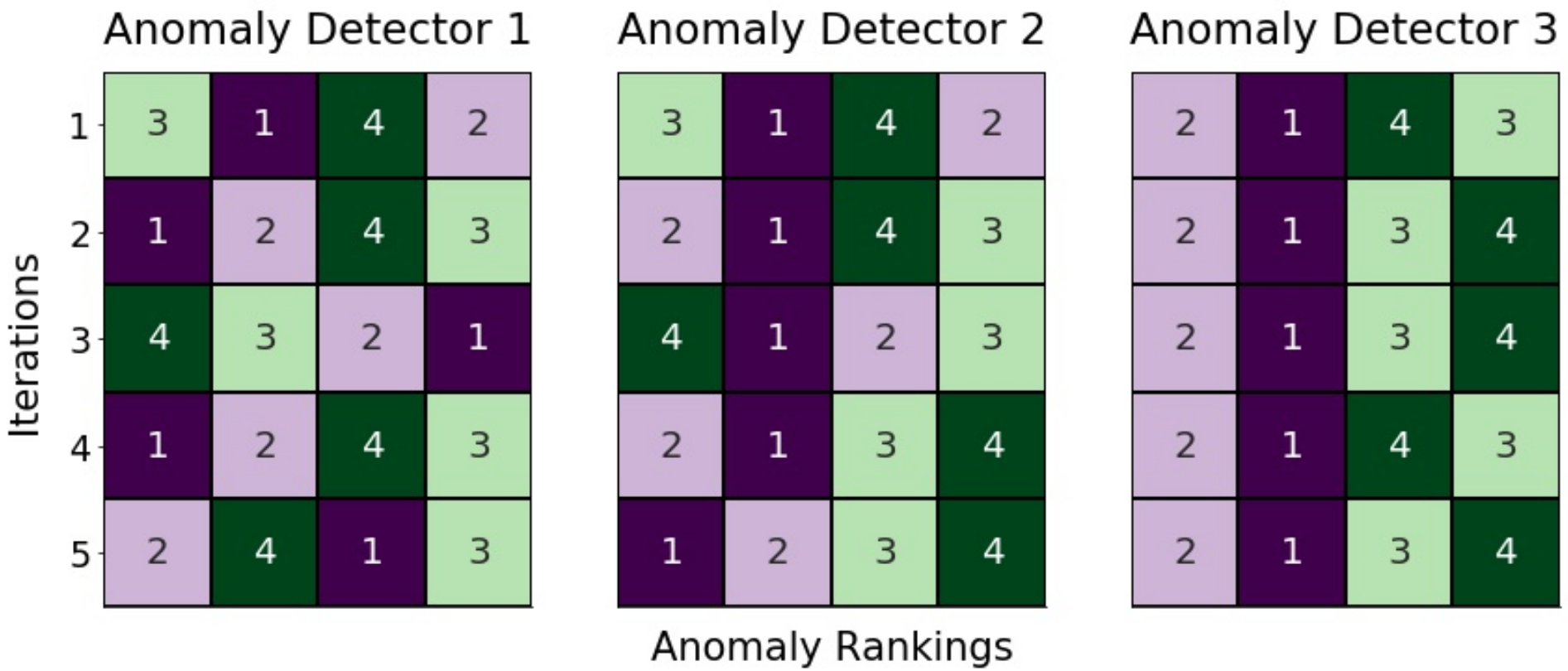
Dataset 3

1	3	1	4	2
2	1	2	4	3
3	4	3	2	1

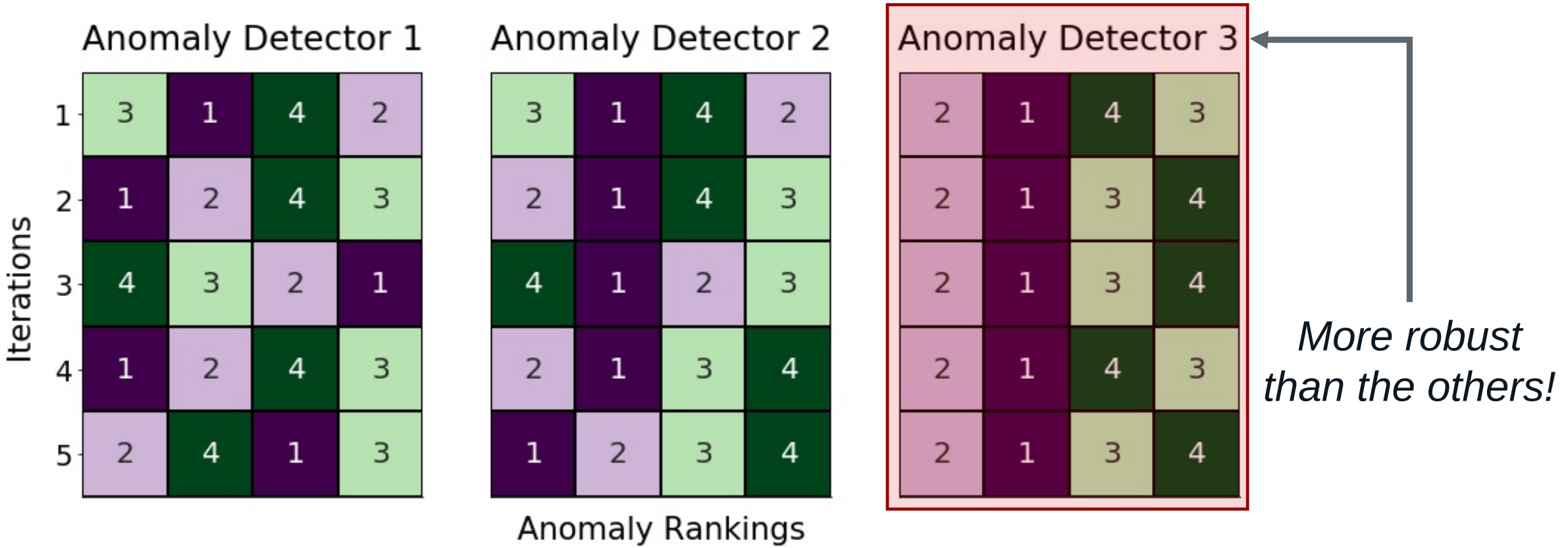
Anomaly rank at each iteration

*How **stable** is the model in making anomaly rankings?*

We Aim To Quantify the Robustness of Anomaly Detectors



We Aim To Quantify the Robustness of Anomaly Detectors



We Aim To Quantify the Robustness of Anomaly Detectors

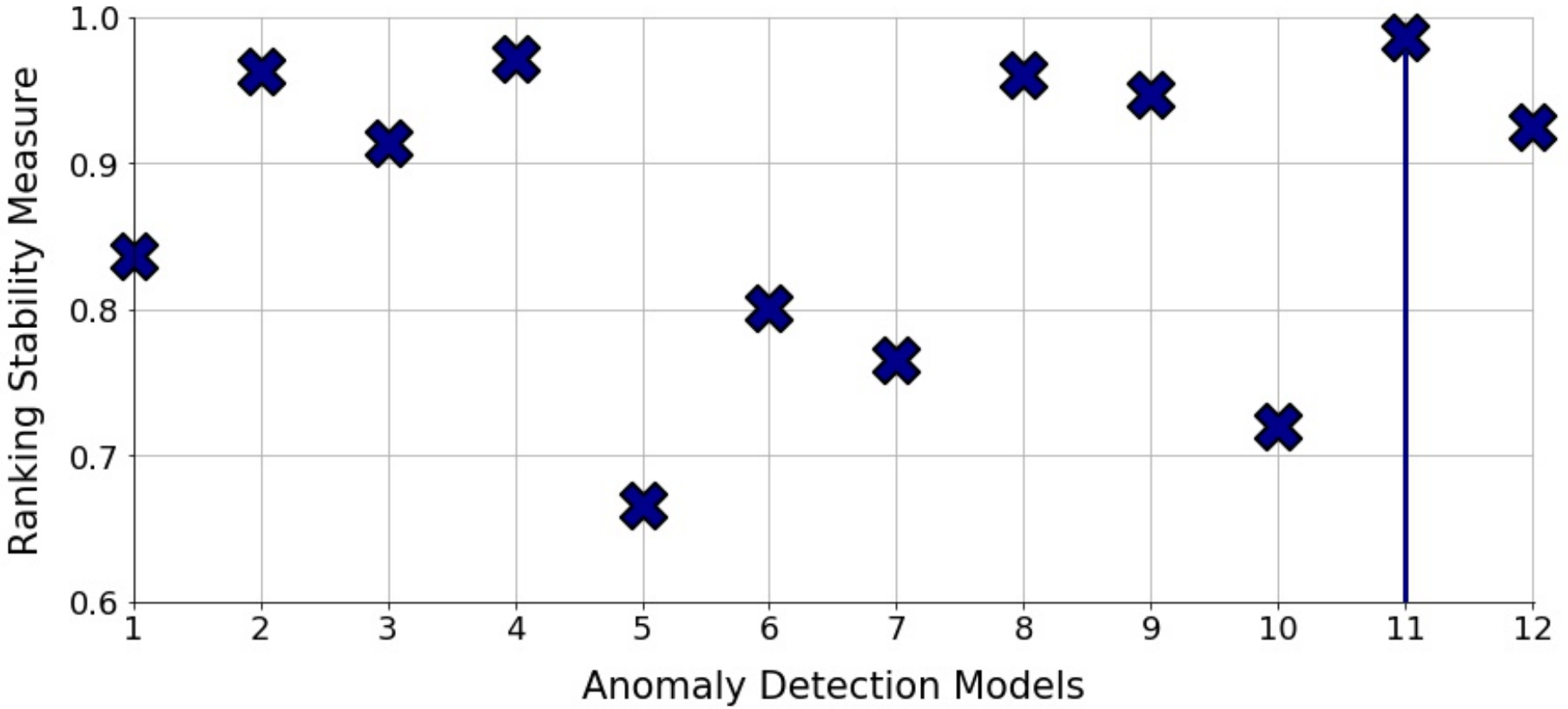
- *Given:*

  Training and test sets, and an anomaly detection model;

- *Do:*

  Design a stability measure quantifying the ability of the model to rank the test examples consistently under variations in the training set.

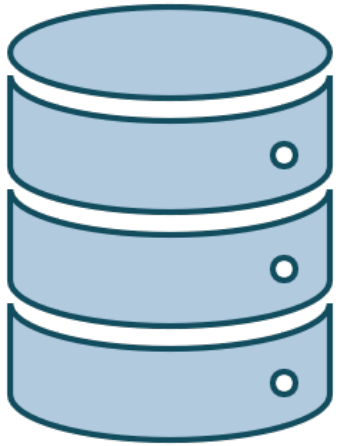
Which Model Would you Trust More?



How to quantify the robustness of any anomaly detector?

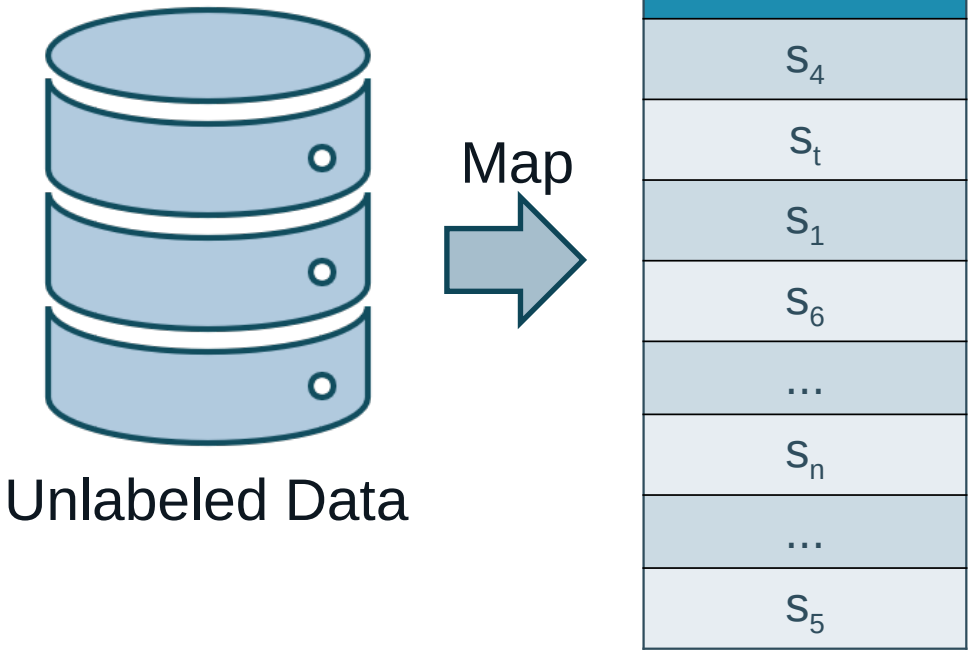
End of the spotlight presentation

3 Steps of Standard Unsupervised Anomaly Detection

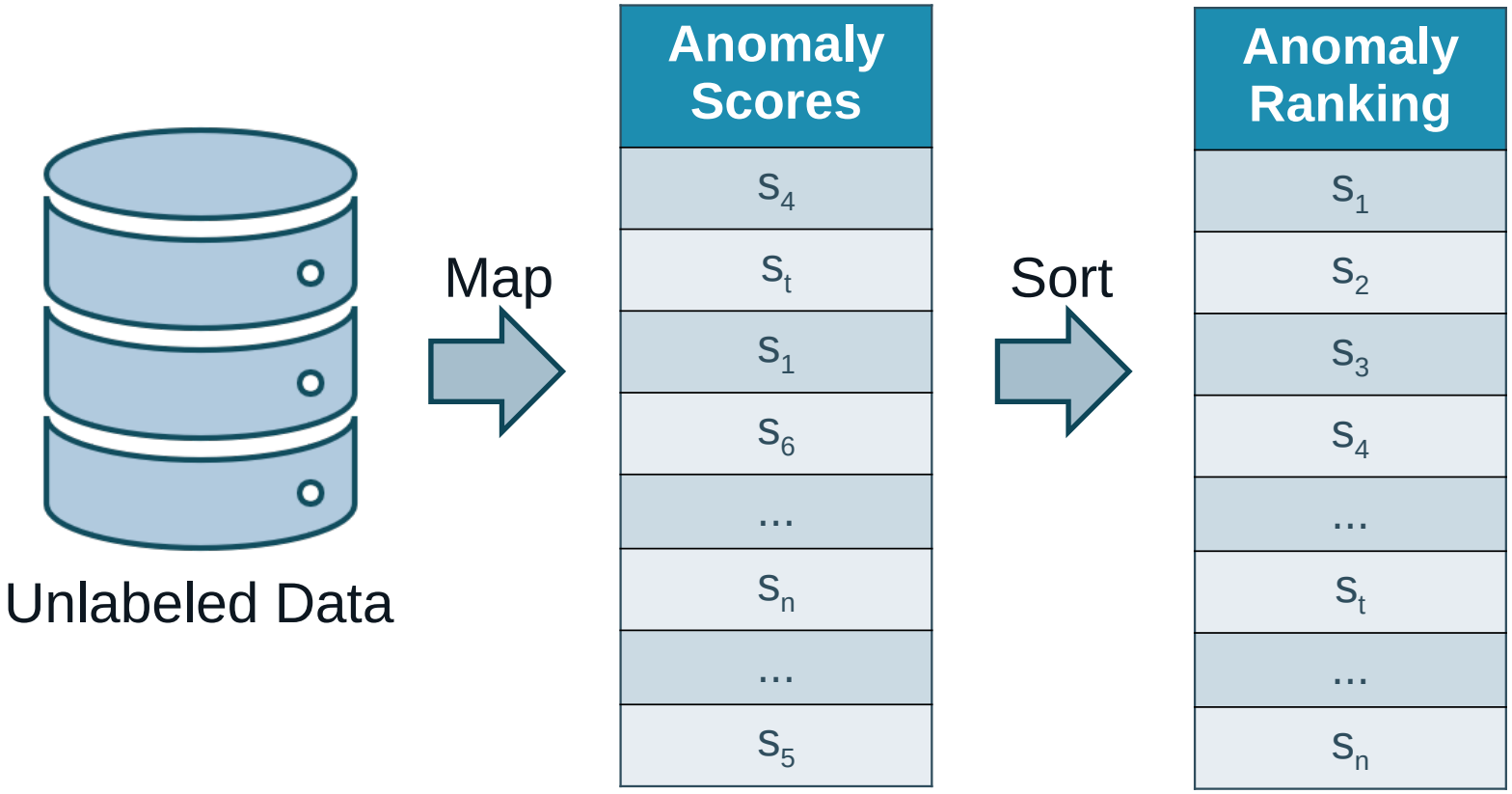


Unlabeled Data

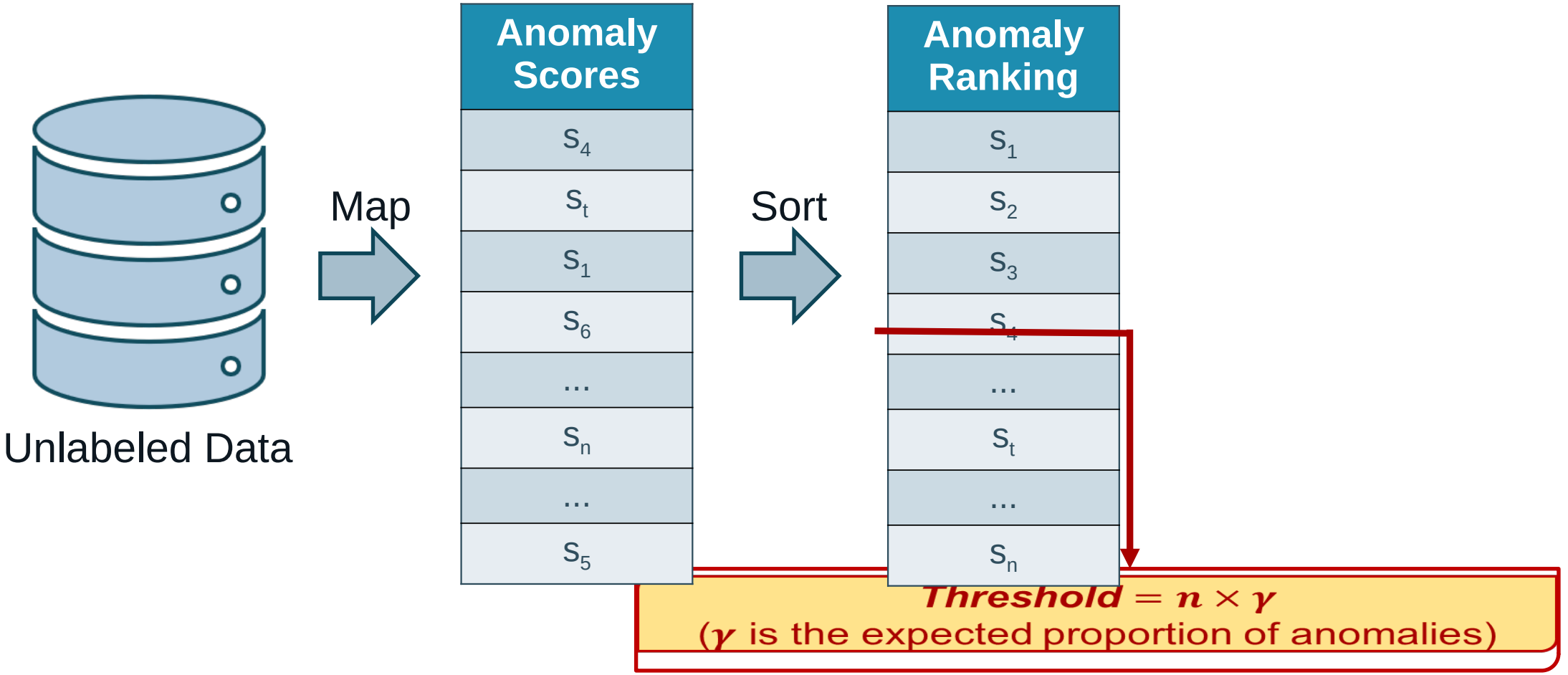
3 Steps of Standard Unsupervised Anomaly Detection



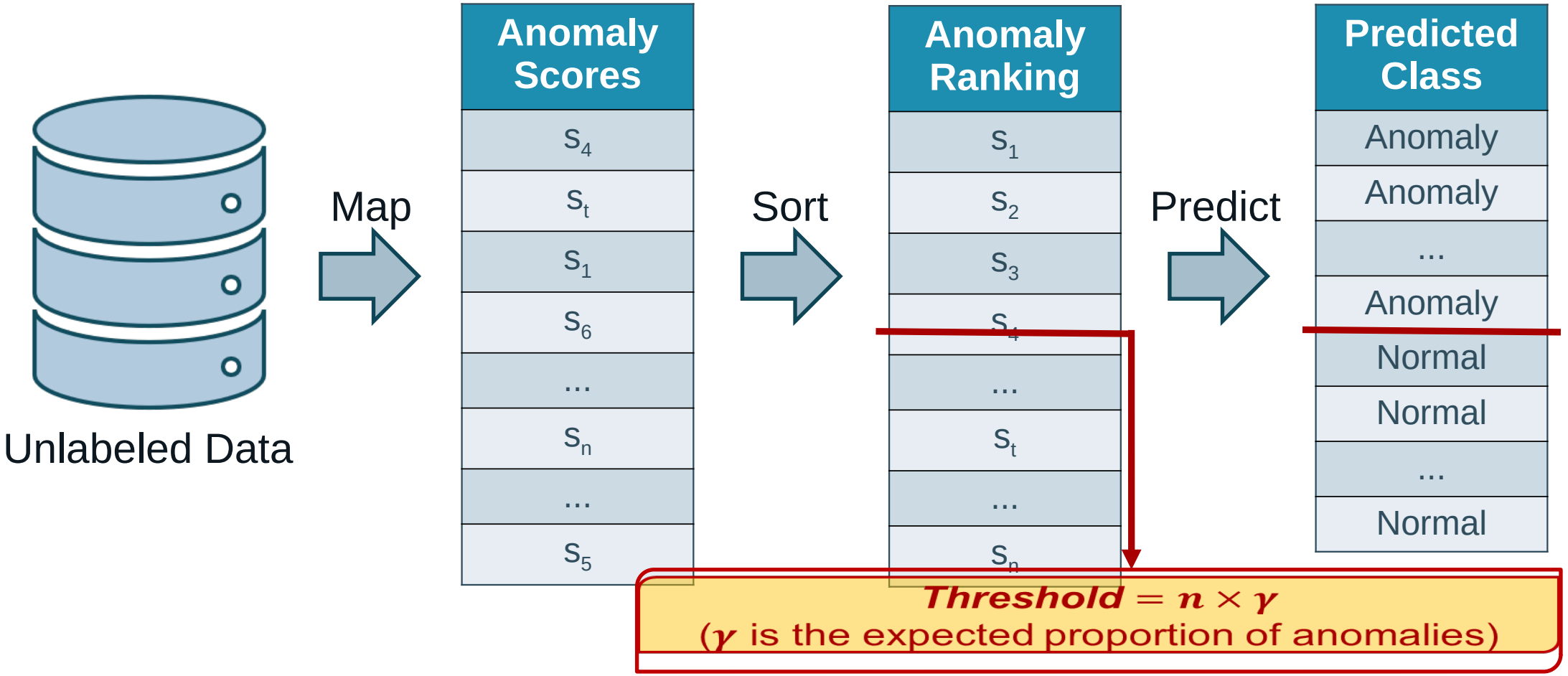
3 Steps of Standard Unsupervised Anomaly Detection



3 Steps of Standard Unsupervised Anomaly Detection



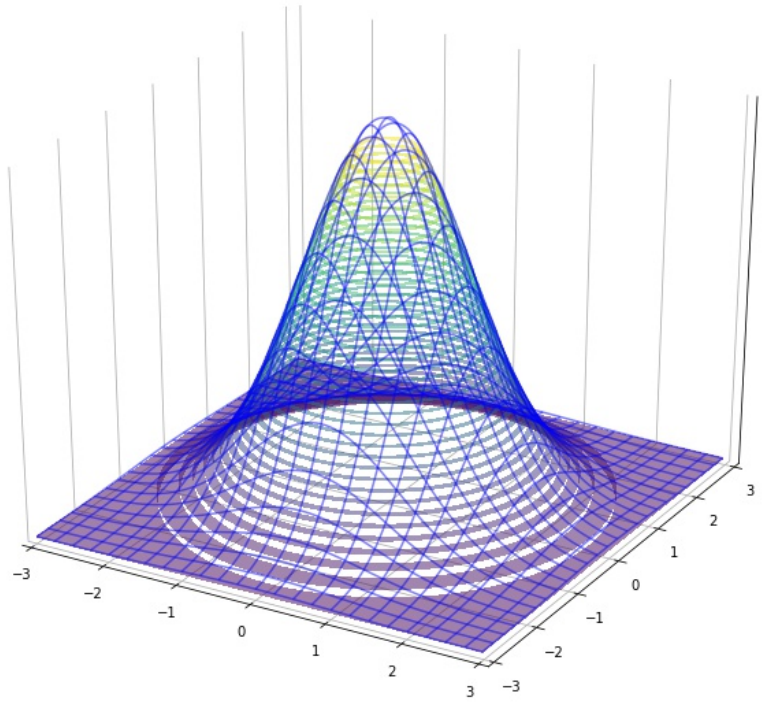
3 Steps of Standard Unsupervised Anomaly Detection



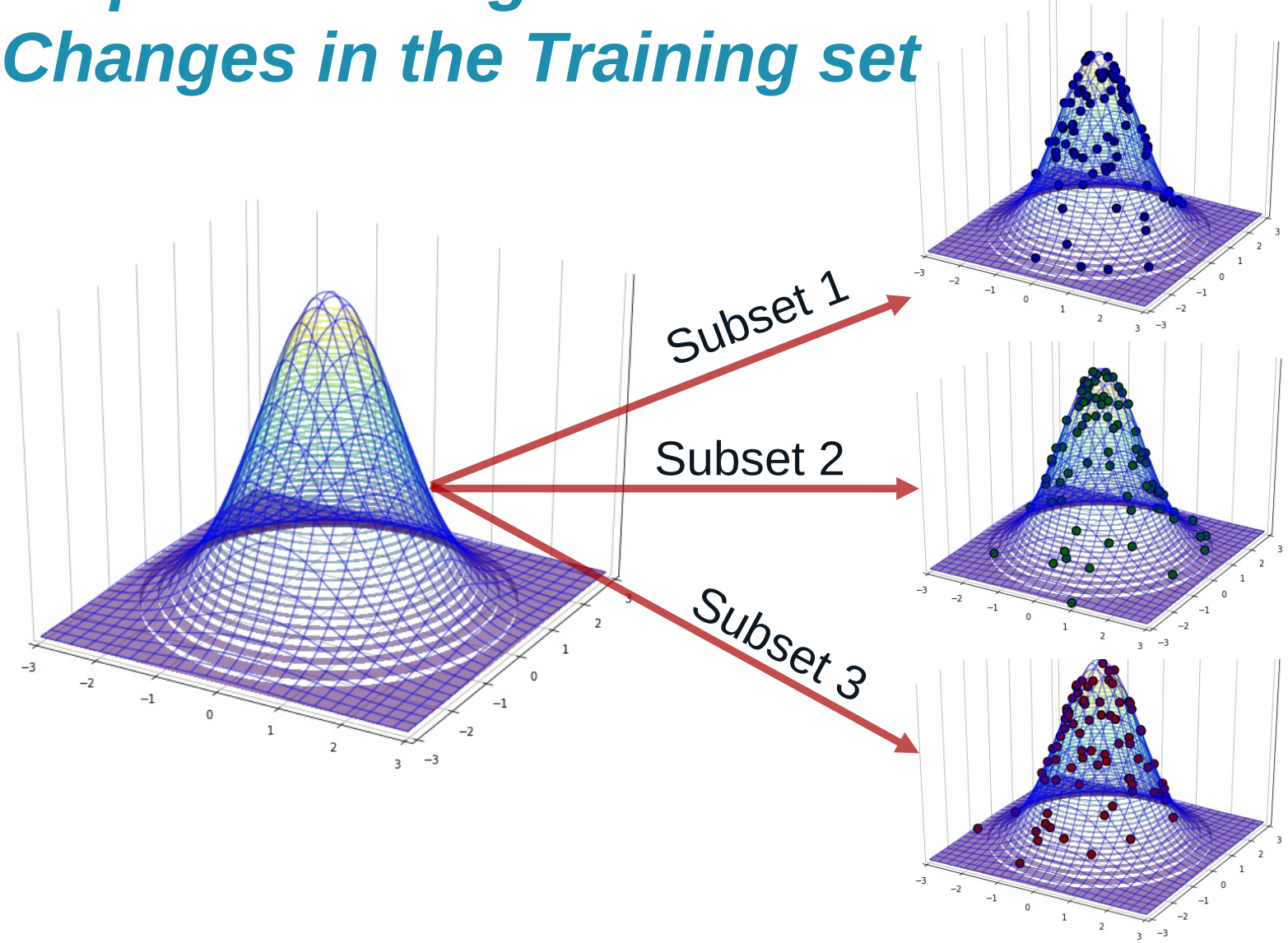
Ranking Stability Measure

A 3-step approach for estimating the robustness

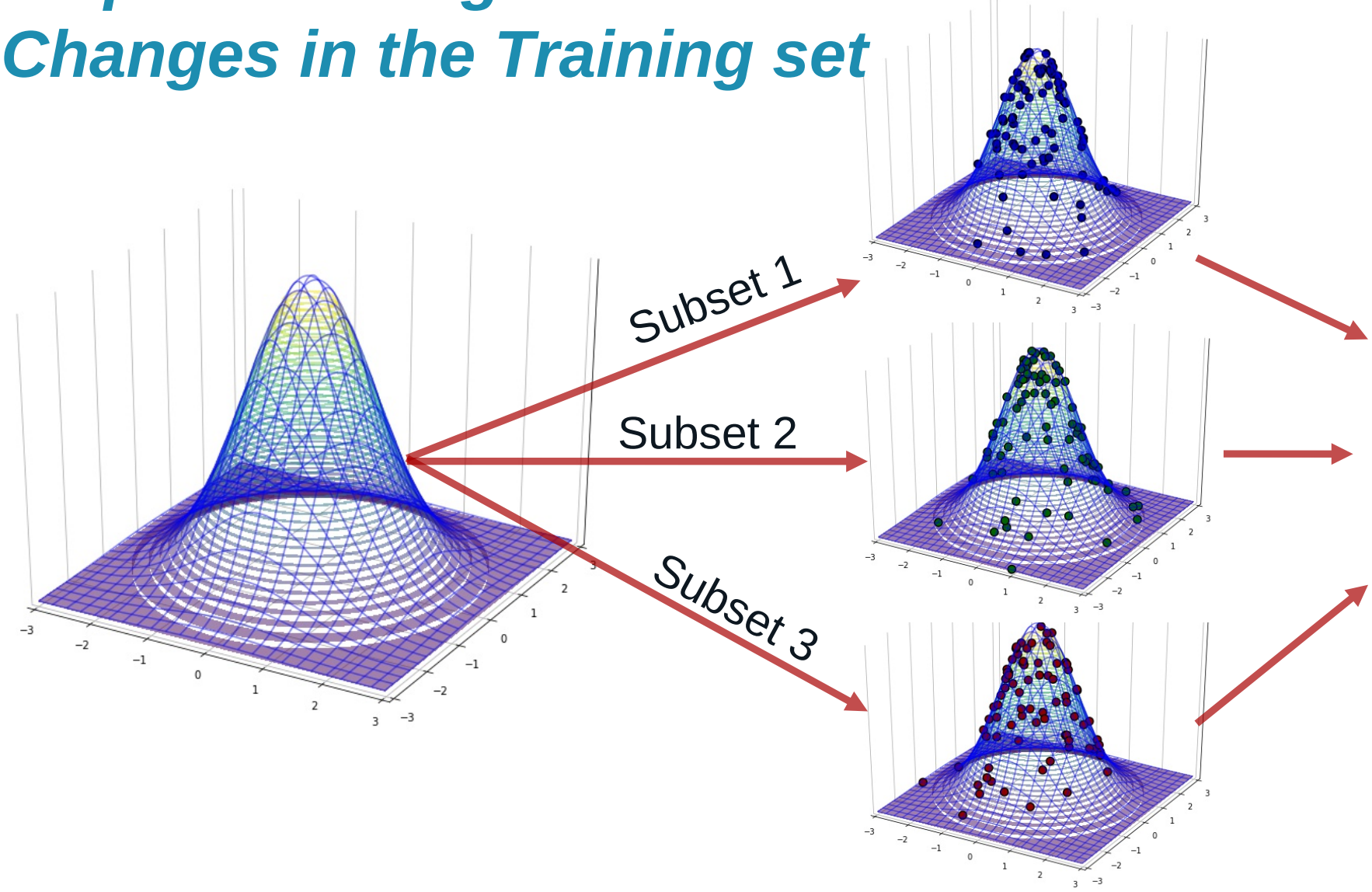
Step 1: Drawing Random Subsets to Simulate Slight Changes in the Training set



Step 1: Drawing Random Subsets to Simulate Slight Changes in the Training set

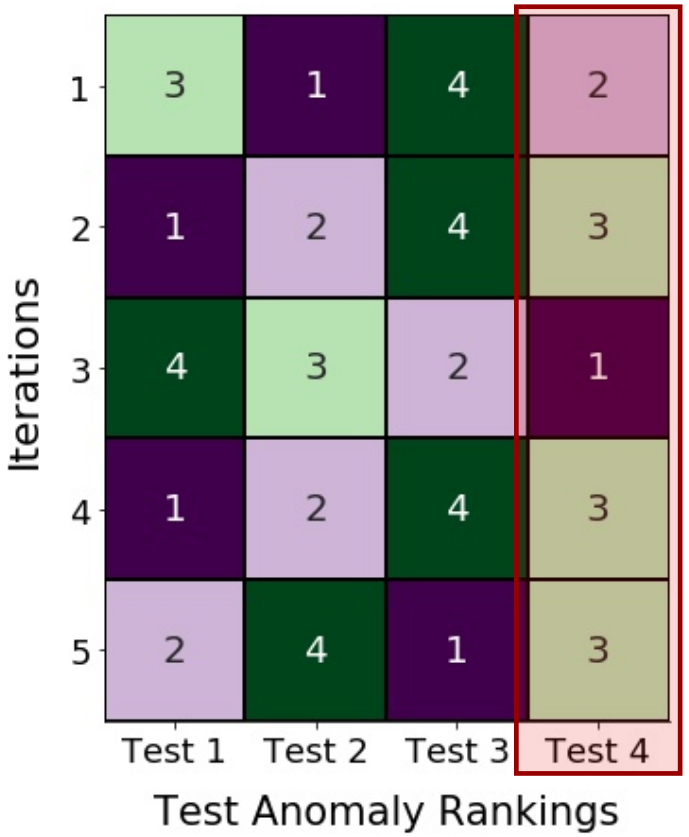


Step 1: Drawing Random Subsets to Simulate Slight Changes in the Training set

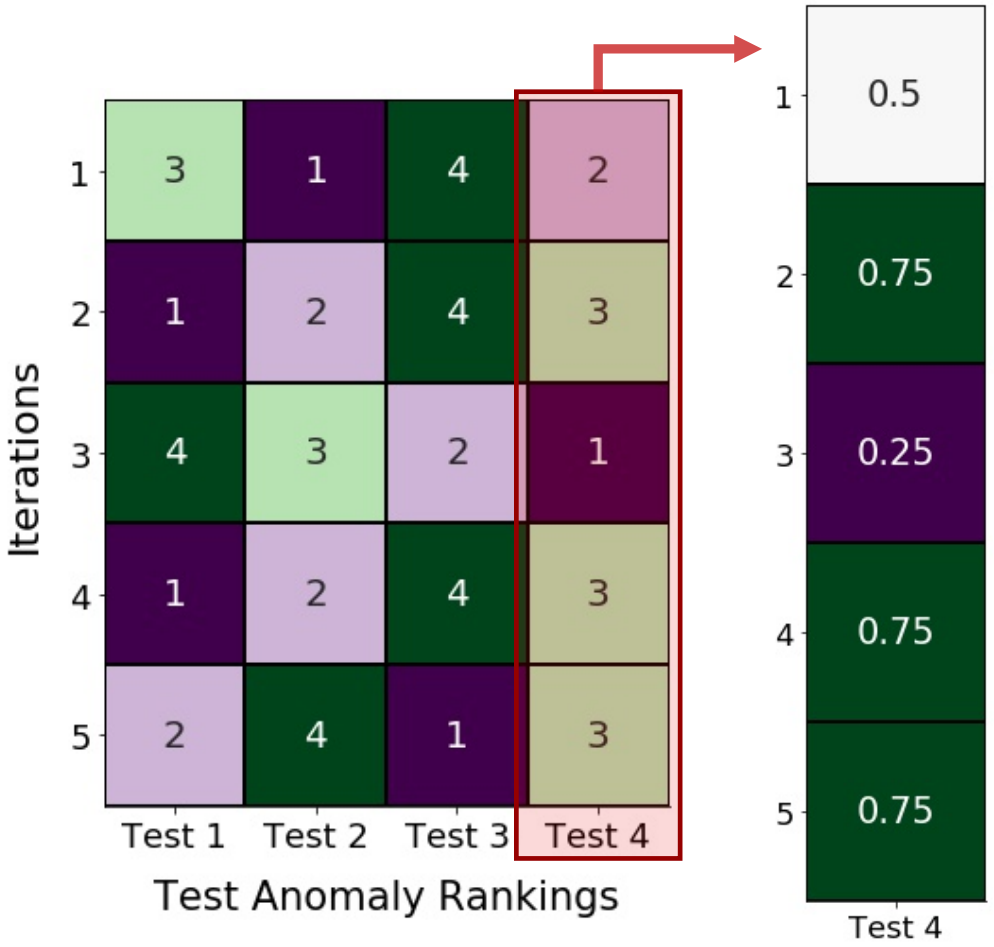


1. Train model
2. Use the model to make the anomaly ranking

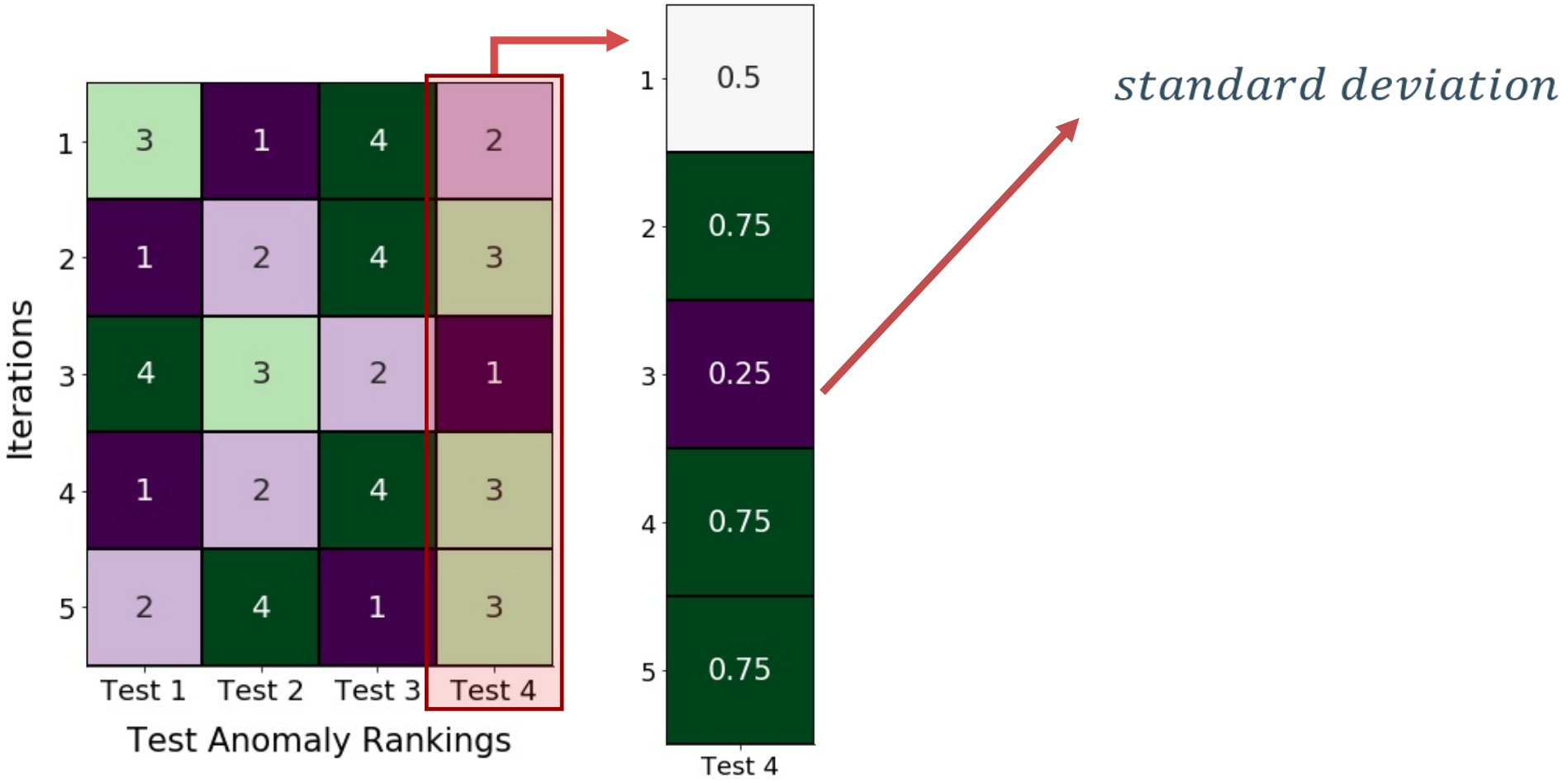
Step 2: Assigning Example-Wise Stability Scores Using the Std. Dev. and the Range of Normalized Positions



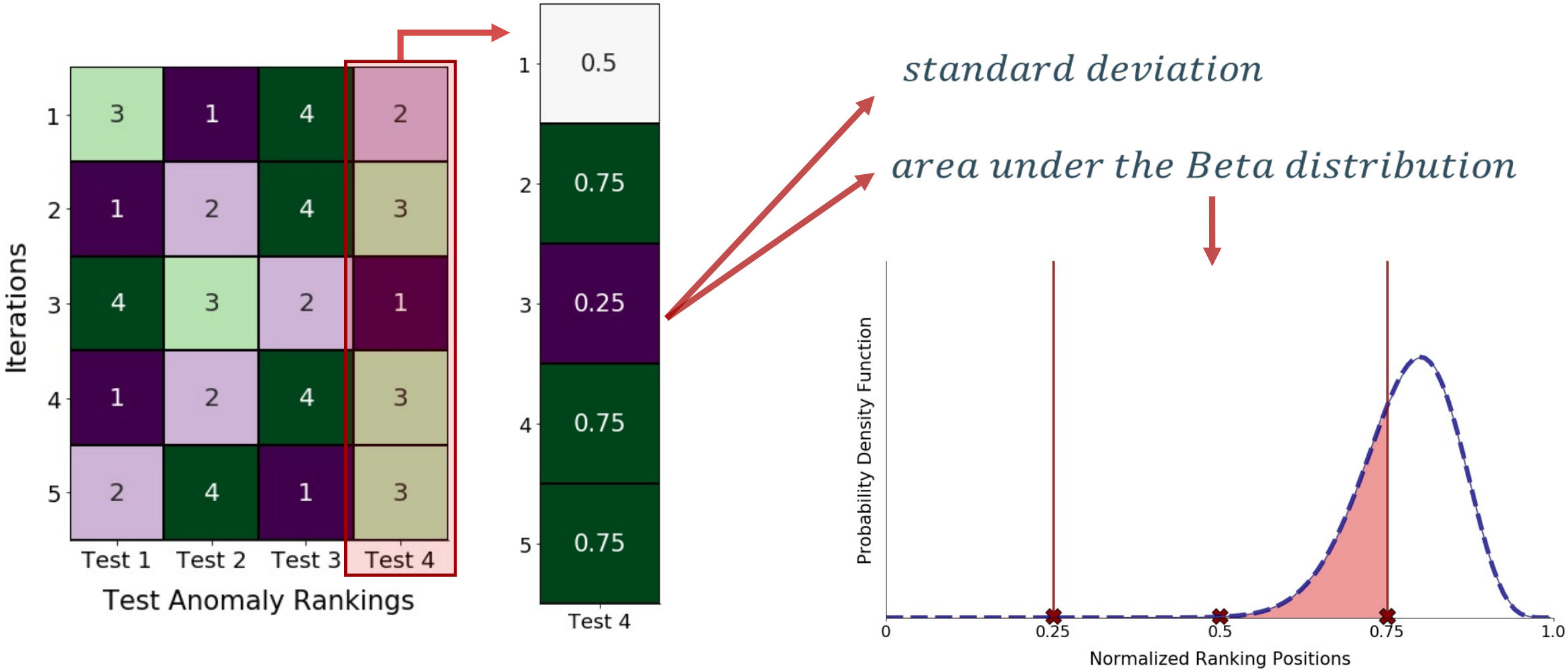
Step 2: Assigning Example-Wise Stability Scores Using the Std. Dev. and the Range of Normalized Positions



Step 2: Assigning Example-Wise Stability Scores Using the Std. Dev. and the Range of Normalized Positions



Step 2: Assigning Example-Wise Stability Scores Using the Std. Dev. and the Range of Normalized Positions



Step 3: Aggregating the Example Scores to a Model Score

Each example has a stability score:

$$Example\ score \approx standard\ deviation \times area\ under\ the\ Beta$$

To obtain the model score, we average the example-wise scores:

$$Model\ stability\ score = \frac{1}{N} \sum_{i=1}^{N} Example\ score\ i$$

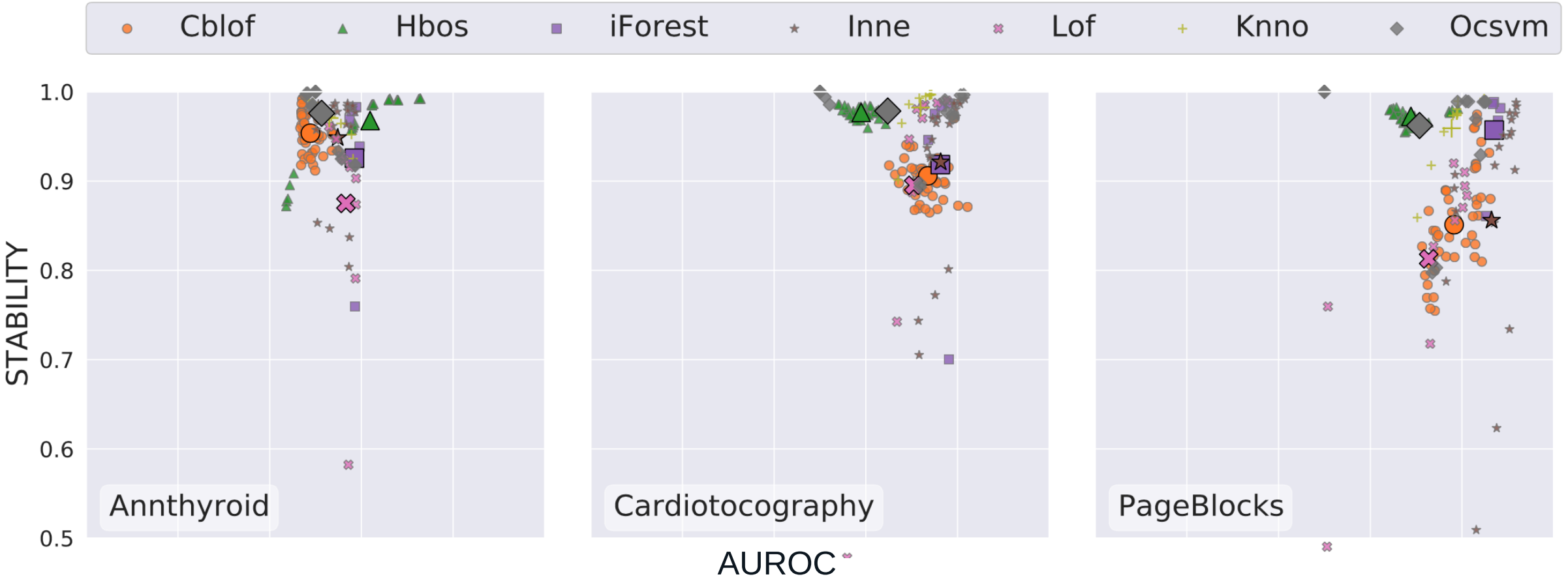
The model score is normalized such that:

- 0 = random rankings;
- 1 = perfect agreement between rankings.

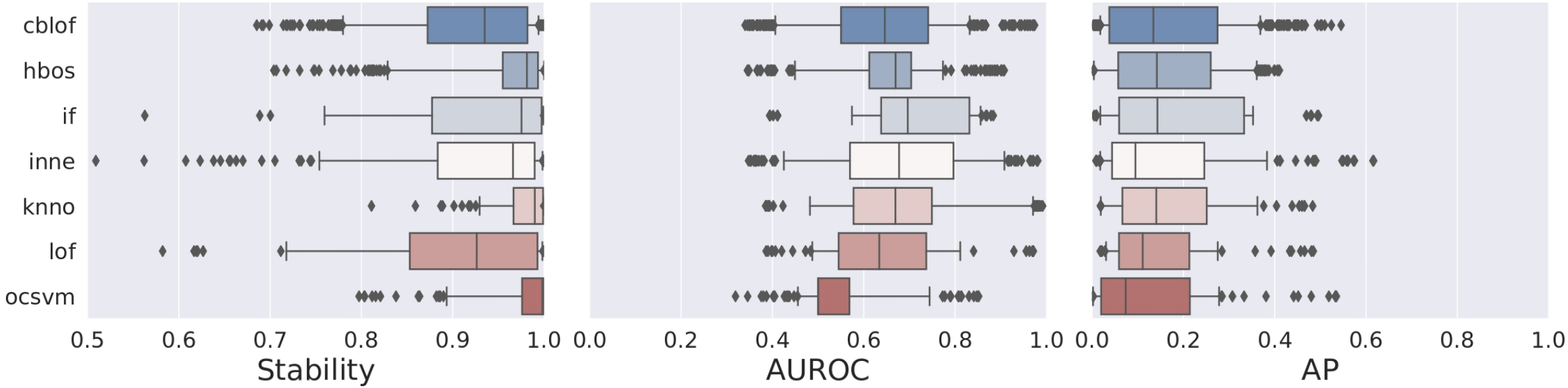
Experiments

Empirical evaluation on benchmark datasets

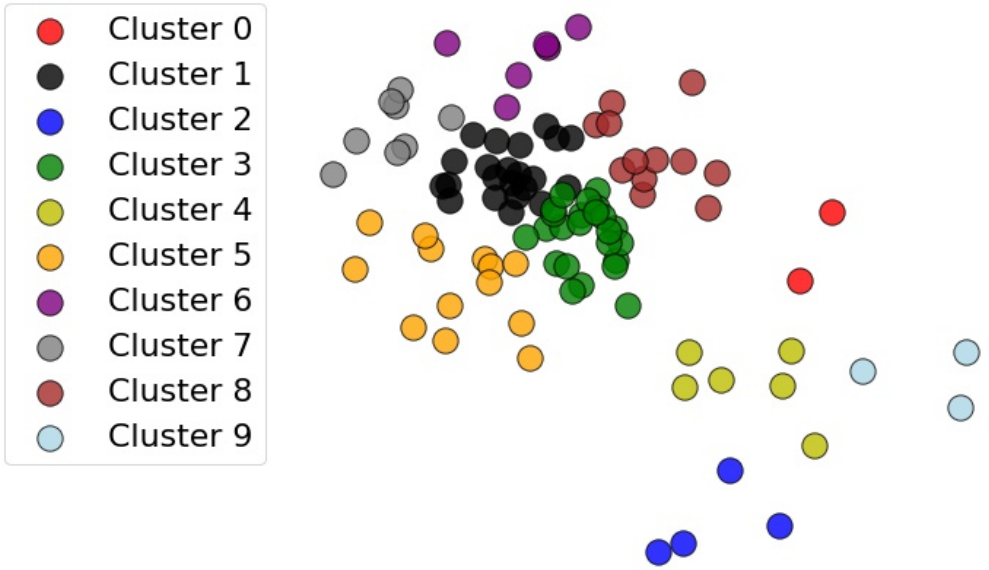
Does Stability Measure Allow Cross-Comparisons Among Anomaly Detectors, Complementing Traditional Measures?



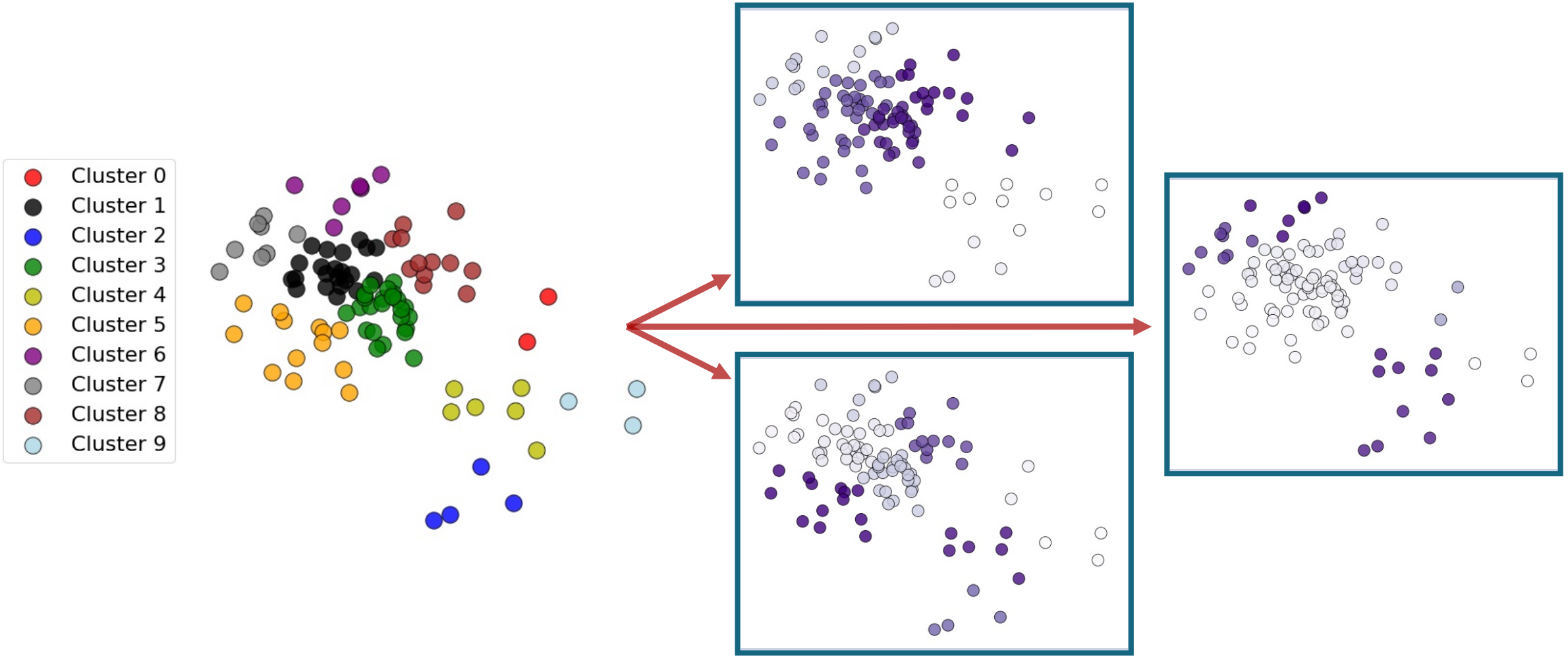
Does Stability Measure Allow Cross-Comparisons Among Anomaly Detectors, Complementing Traditional Measures?



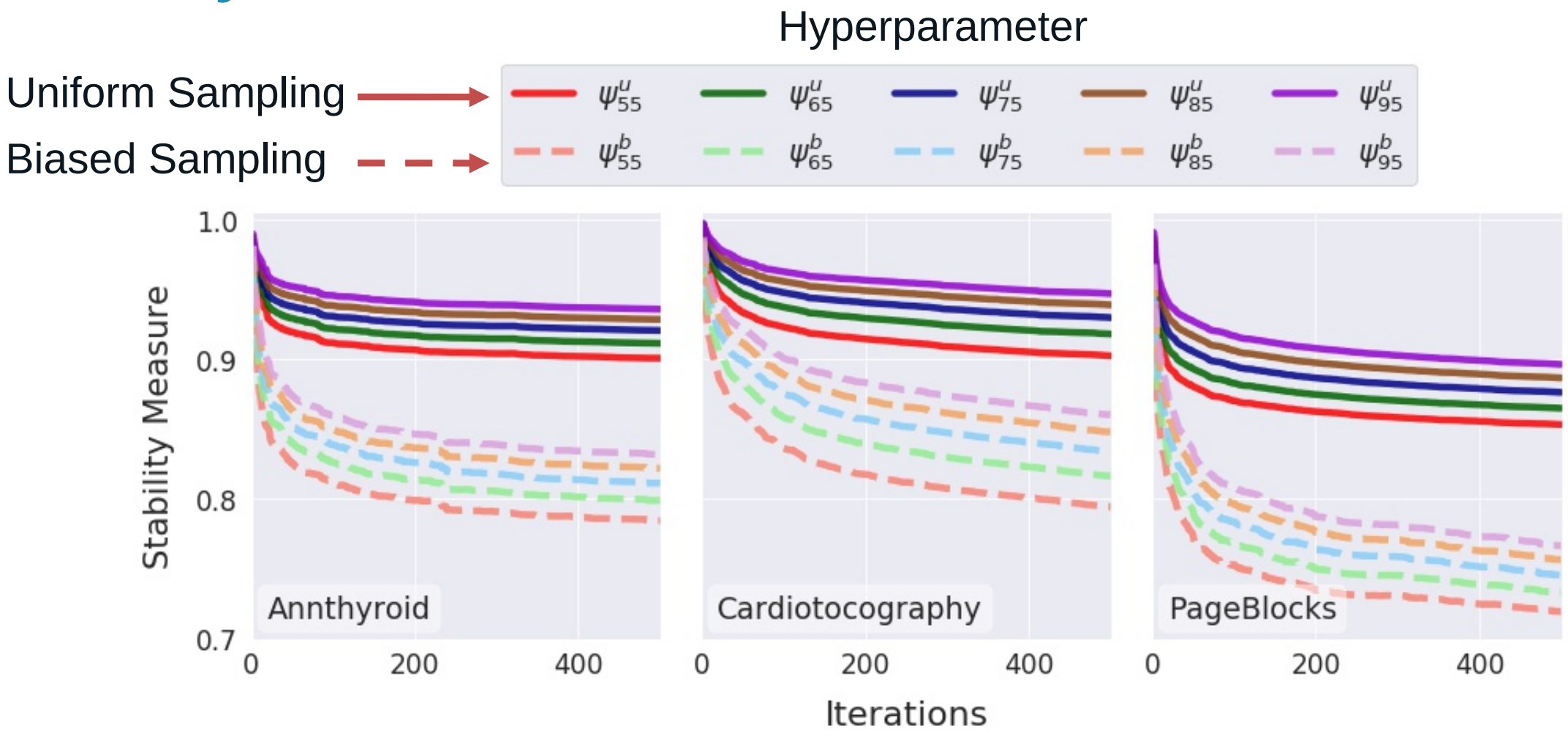
How does our Stability Measure behave when Biased Subsets are drawn?



How does our Stability Measure behave when Biased Subsets are drawn?



Biased Subsampling Strategies Lead to Less Stable Anomaly Detectors



Stability Measure Is a Valid Alternative To Traditional Performance Measures to Quantify Models' Behaviour

We proposed a ***novel stability measure*** to quantify the robustness of anomaly detection methods:

1. This measure allows meaningful cross-comparison of different methods;

2. This measure behaves realistically in our experiments.

All code and experiments are available online:

https://github.com/Lorenzo-Perini/StabilityRankings_AD

KU LEUVEN DTAI Declarative Languages & Artificial Intelligence

Contact us

name.surname@kuleuven.be

Lorenzo Perini, *Connor Galvin, Vincent Vercruyssen*

https://people.cs.kuleuven.be/~lorenzo.perini/

@LorenzoPerini95

EDML 2020