

ACTIVE LEARNING MODAL

PRESENTED TO
Eng.Salah ElAbyad

PRESENTED BY
Yomna Taher Abd allah
Adel Abd Elmoneam

22/4/2023

ACTIVE LEARNING

Is "a method of learning in which students are actively or experientially involved in the learning process and where there are different levels of active learning, depending on student involvement.

Strategies

1

Margin Sampling

3

Random sampling

2

The Query by Committee

4

Entropy Sampling

5

Uncertainty sampling

1

Margin Sampling

The shortcoming of the LC strategy is that it only takes into consideration the most probable label and disregards the other label probabilities. The margin sampling strategy seeks to overcome this disadvantage by selecting the instance that has the smallest difference between the first and second most probable labels.

2

The Query by Committee

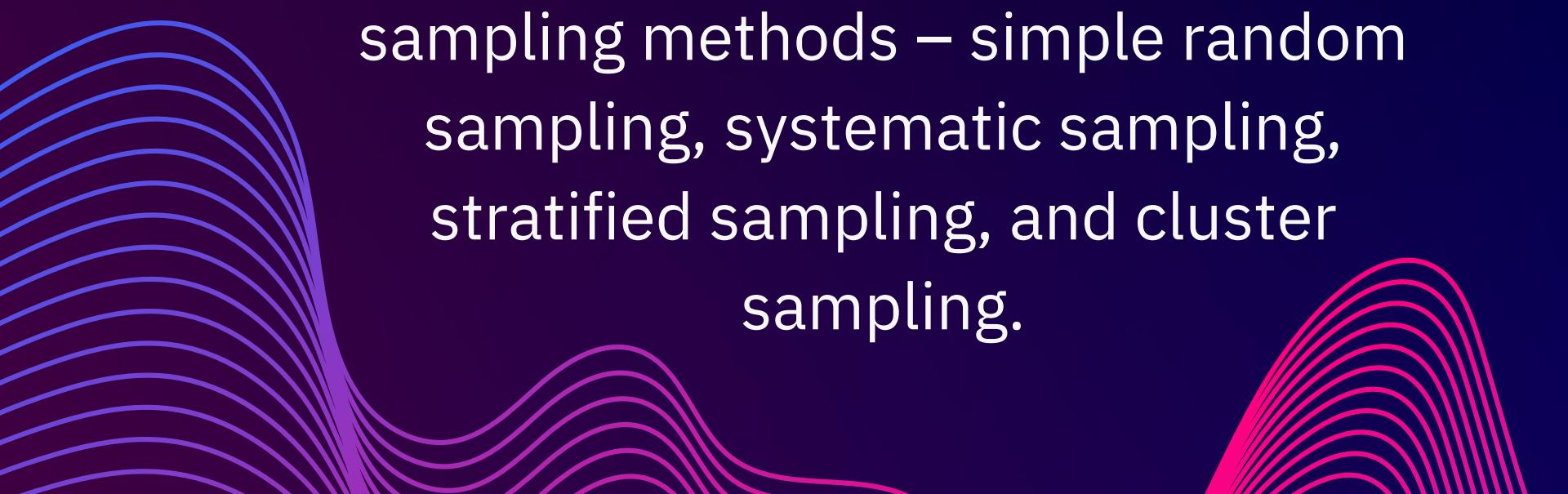
Algorithm (QBC) was the inspiration behind many algorithms in the selective sampling framework [3, 4, 5]. QBC is a simple yet powerful algorithm. During learning it maintains a version space, the space of all the classifiers which are consistent with all the previous labeled instances. Whenever an unlabeled instance is available, QBC selects two random hypotheses from the version space and only queries for the label of the new instance if the two hypotheses disagree.



3

Random sampling

Also known as probability sampling, is a sampling method that allows for the randomization of sample selection. It is essential to keep in mind that samples do not always produce an accurate representation of a population in its entirety; hence, any variations are referred to as sampling errors. There are four primary, random (probability) sampling methods – simple random sampling, systematic sampling, stratified sampling, and cluster sampling.



4

Entropy Sampling

This strategy measures the uncertainty of the predicted probability distribution of the classes.

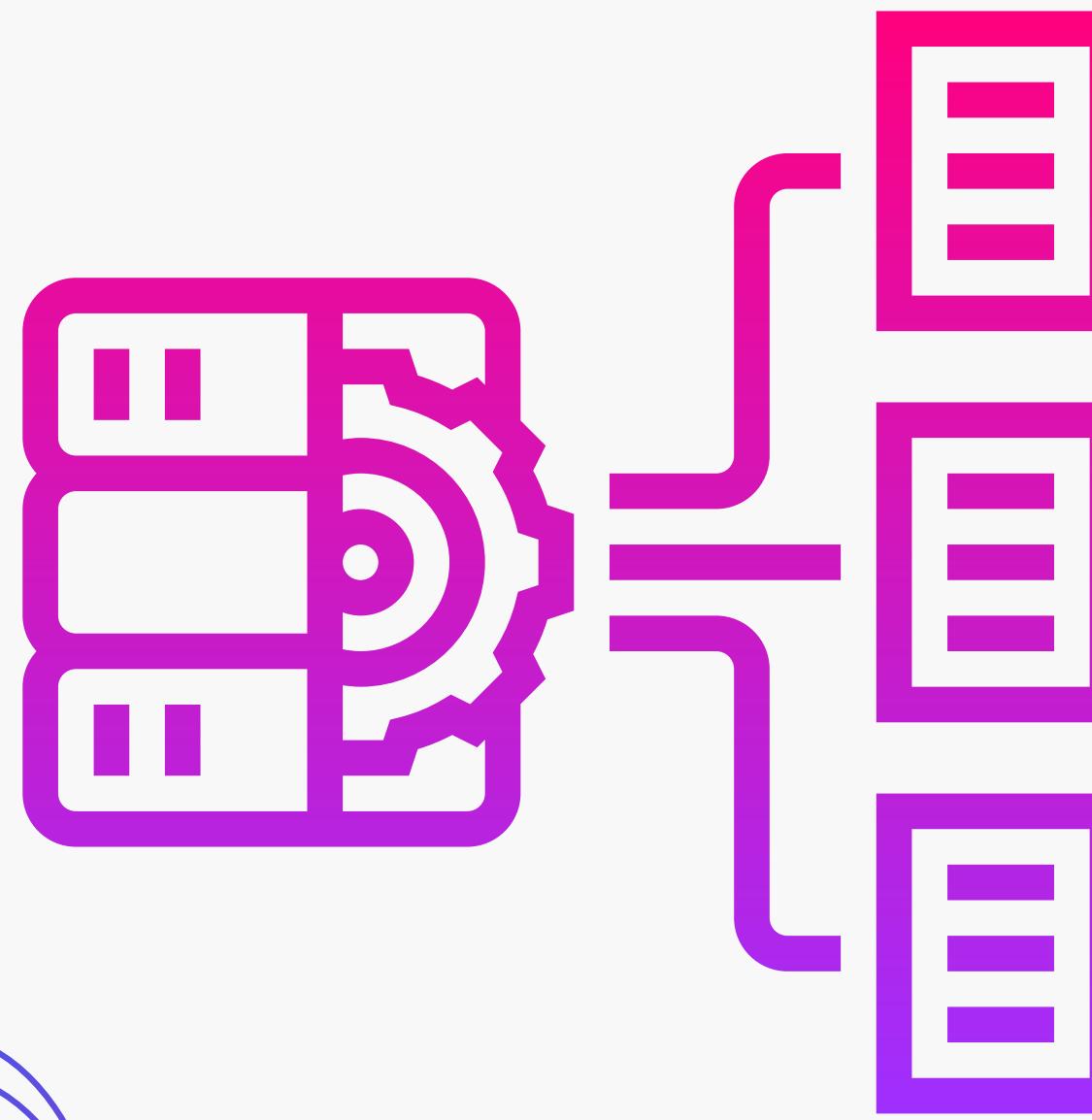
In other words, the query strategy chooses samples for which the model is most uncertain about the predicted probability distribution of the classes.

In the code, it is implemented by selecting the samples with the highest entropy as measured by the predicted probability distribution.

This strategy is useful when the model has low confidence or exhibits poor performance.

Uncertainty sampling

Which is among the most popular approaches, the active learner sequentially queries the label of those instances for which its current prediction is maximally uncertain. Predictions as well as the measures used to quantify the degree of uncertainty, such as entropy, are traditionally of a probabilistic nature. Yet, alternative approaches to capturing uncertainty in machine learning, alongside with corresponding uncertainty measures, have been proposed in recent years. Some of these measures seek to distinguish different sources and to separate different types of uncertainty, such as the reducible (epistemic) and the irreducible (aleatoric) part of the total uncertainty in a prediction. The goal of this paper is to elaborate on the usefulness of such measures for uncertainty sampling, and to compare their performance in active learning.



1.Iris dataset

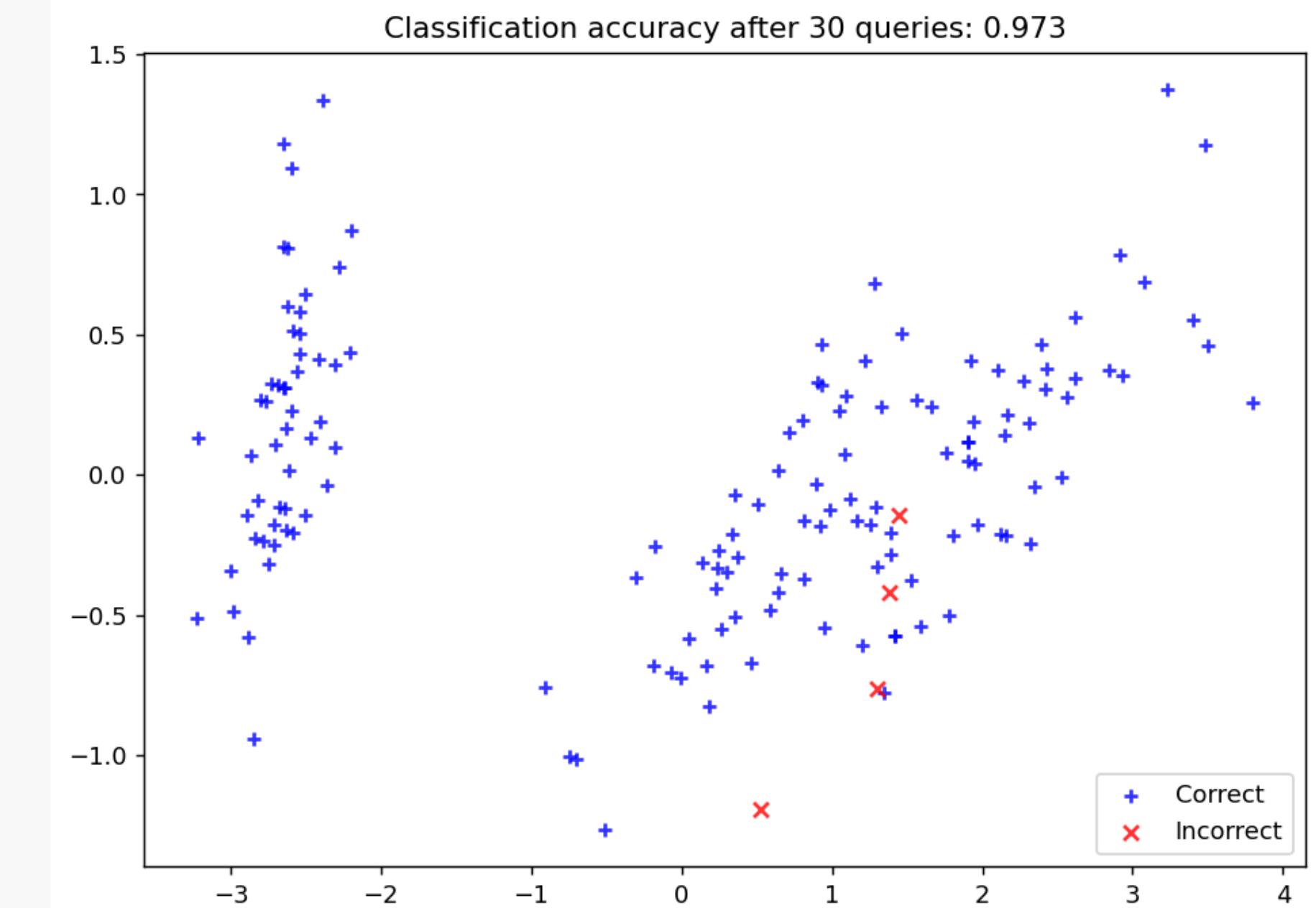
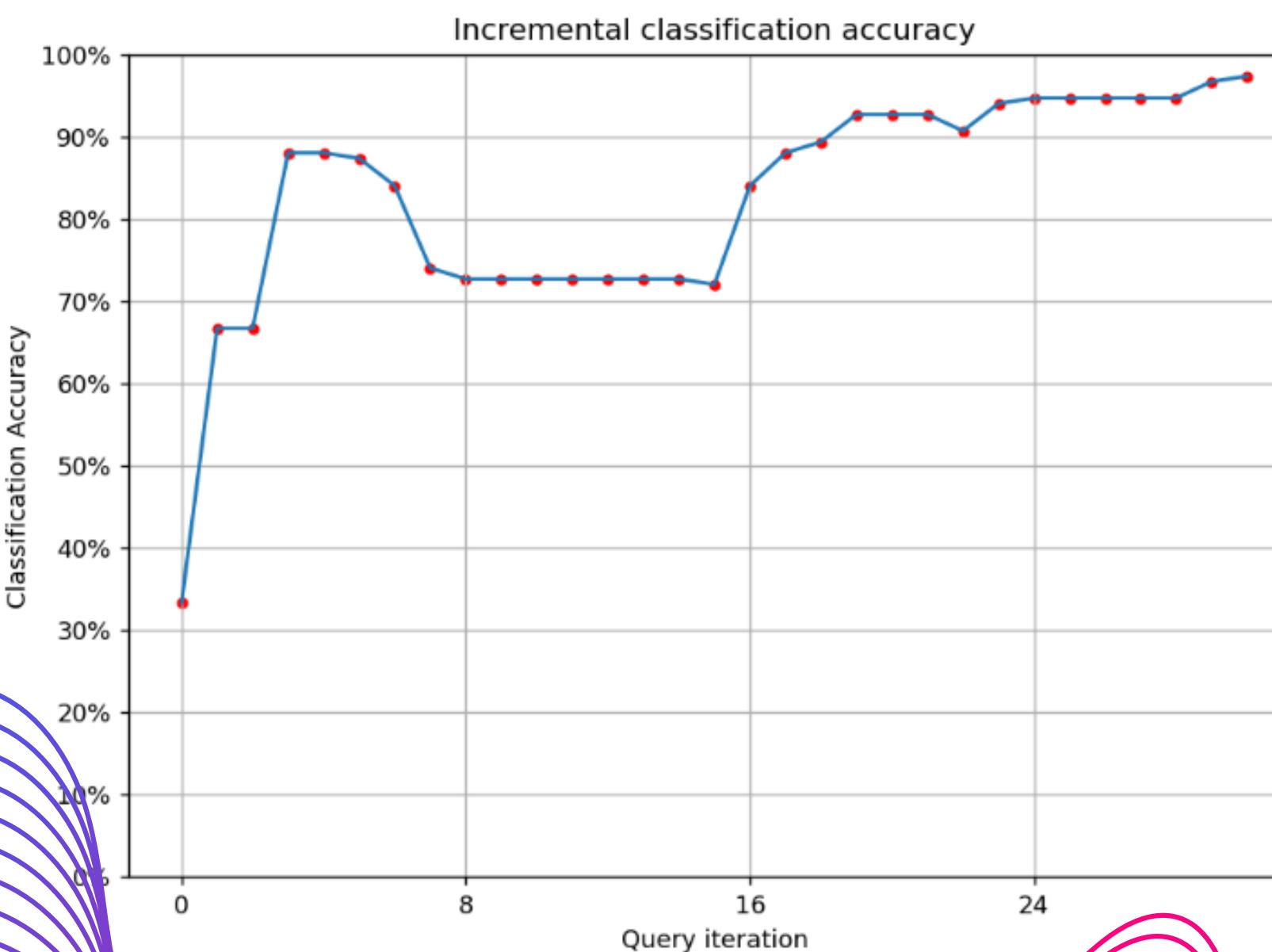
2.Corona_NLP dataset

3.heart dataset

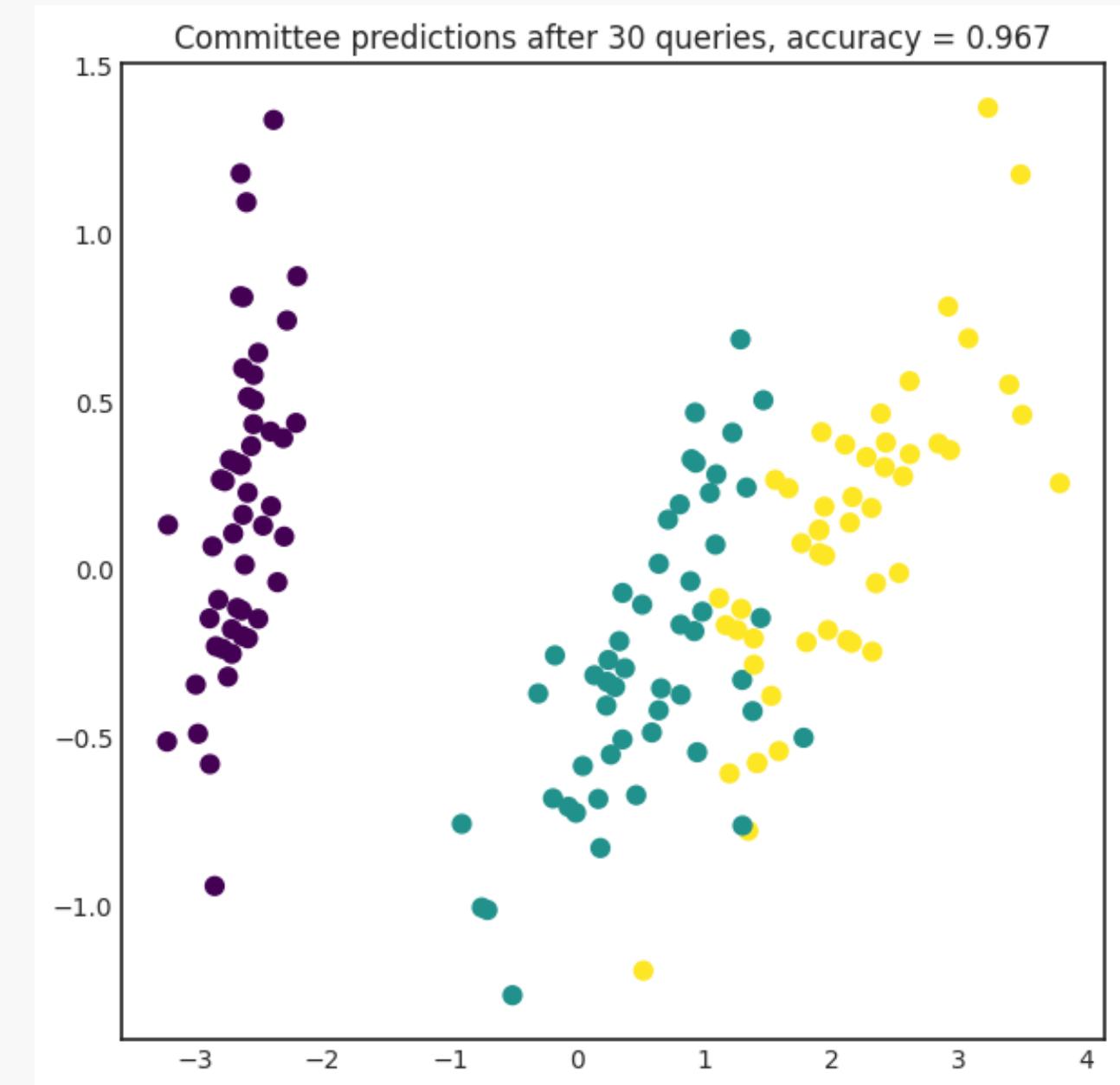
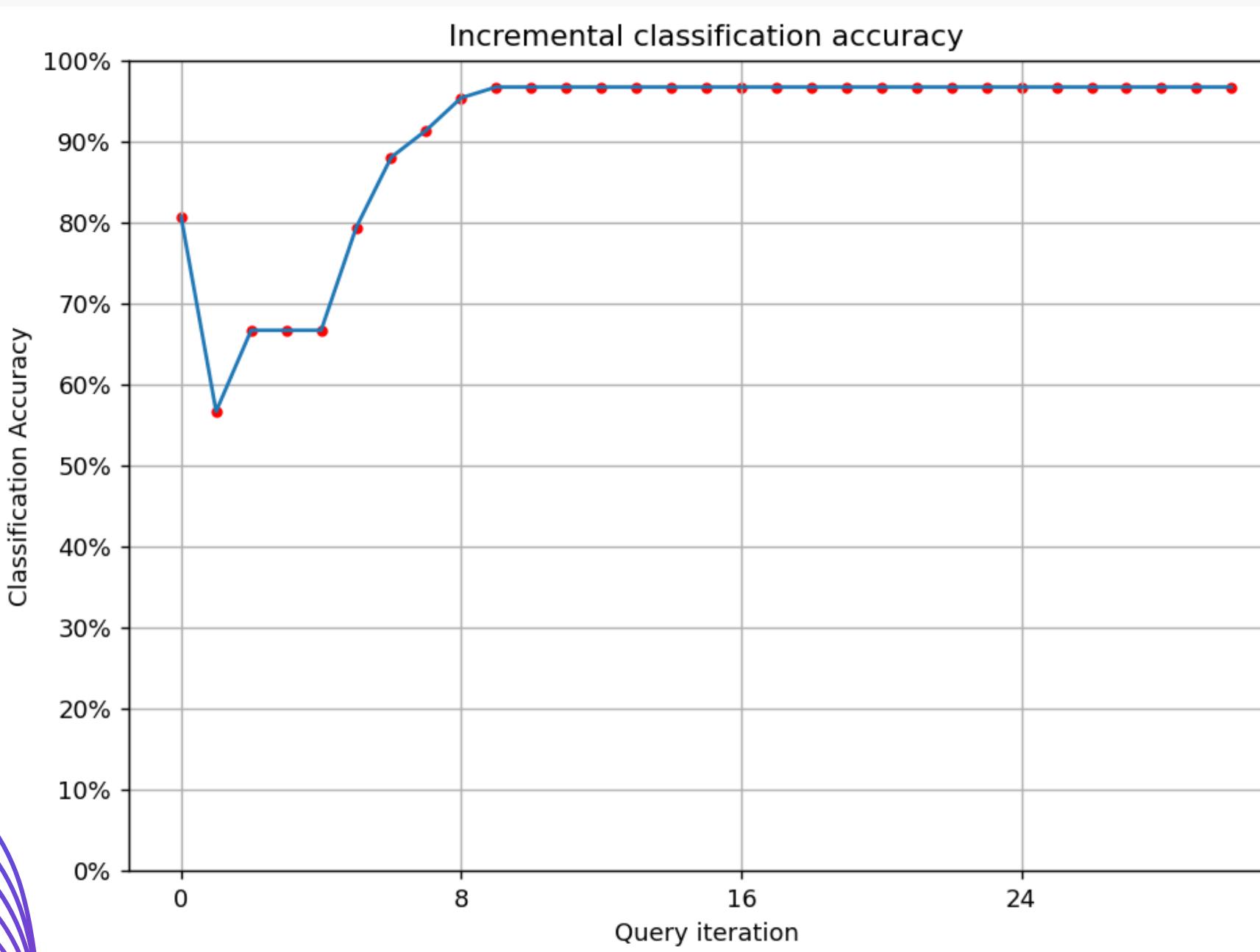
1

Iris dataset

1

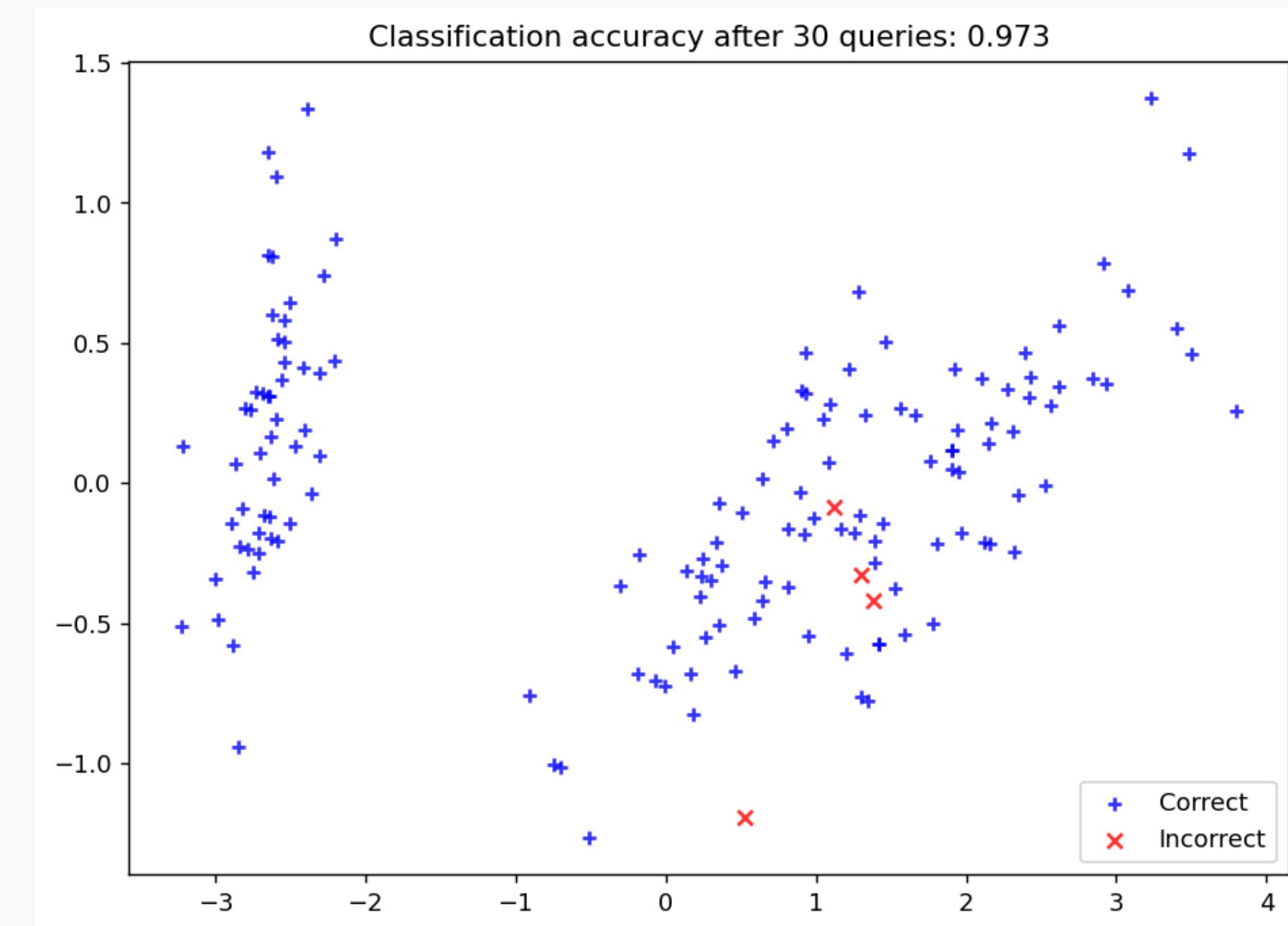
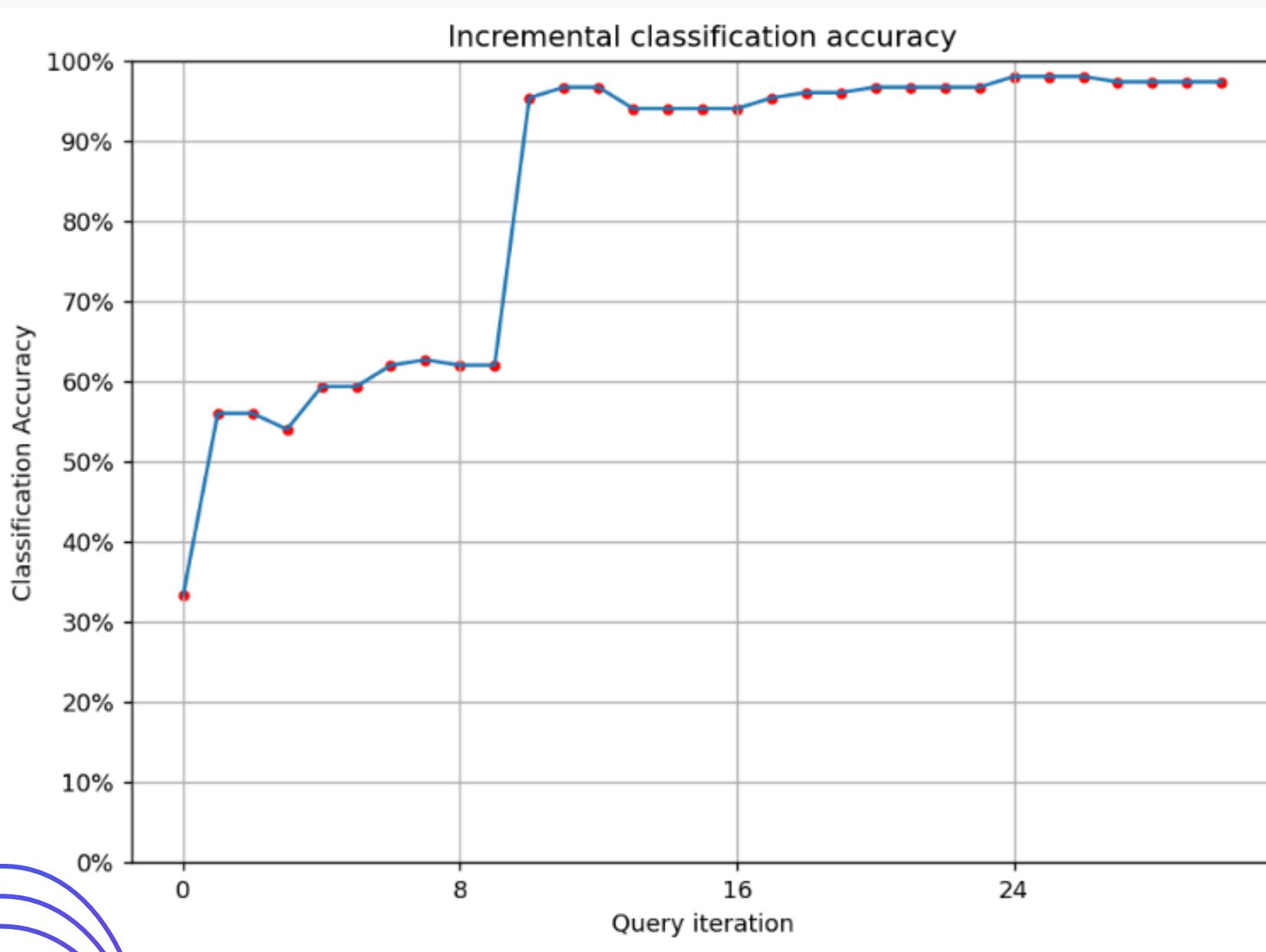
Margin Sampling

2

The Query by Committee

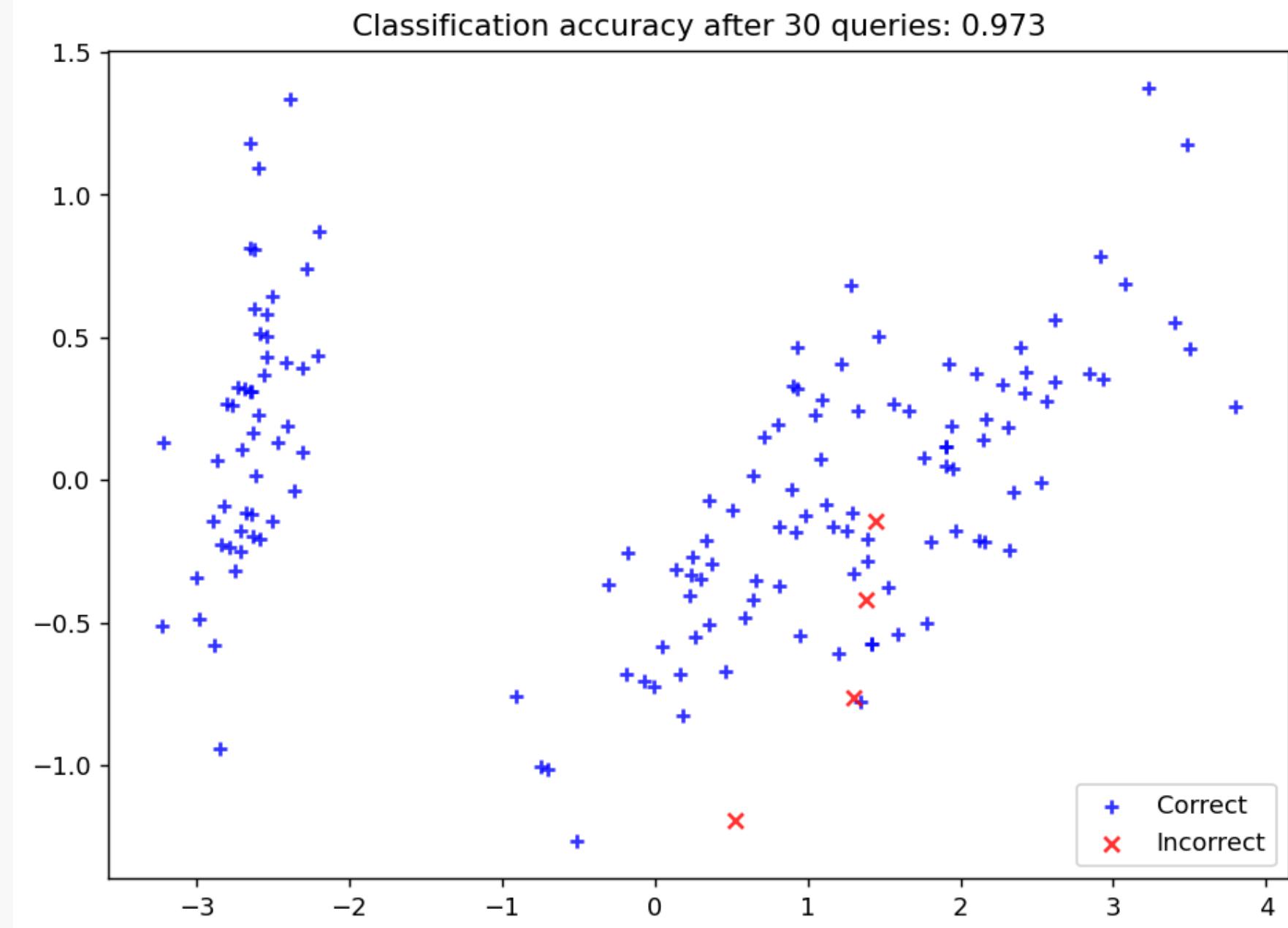
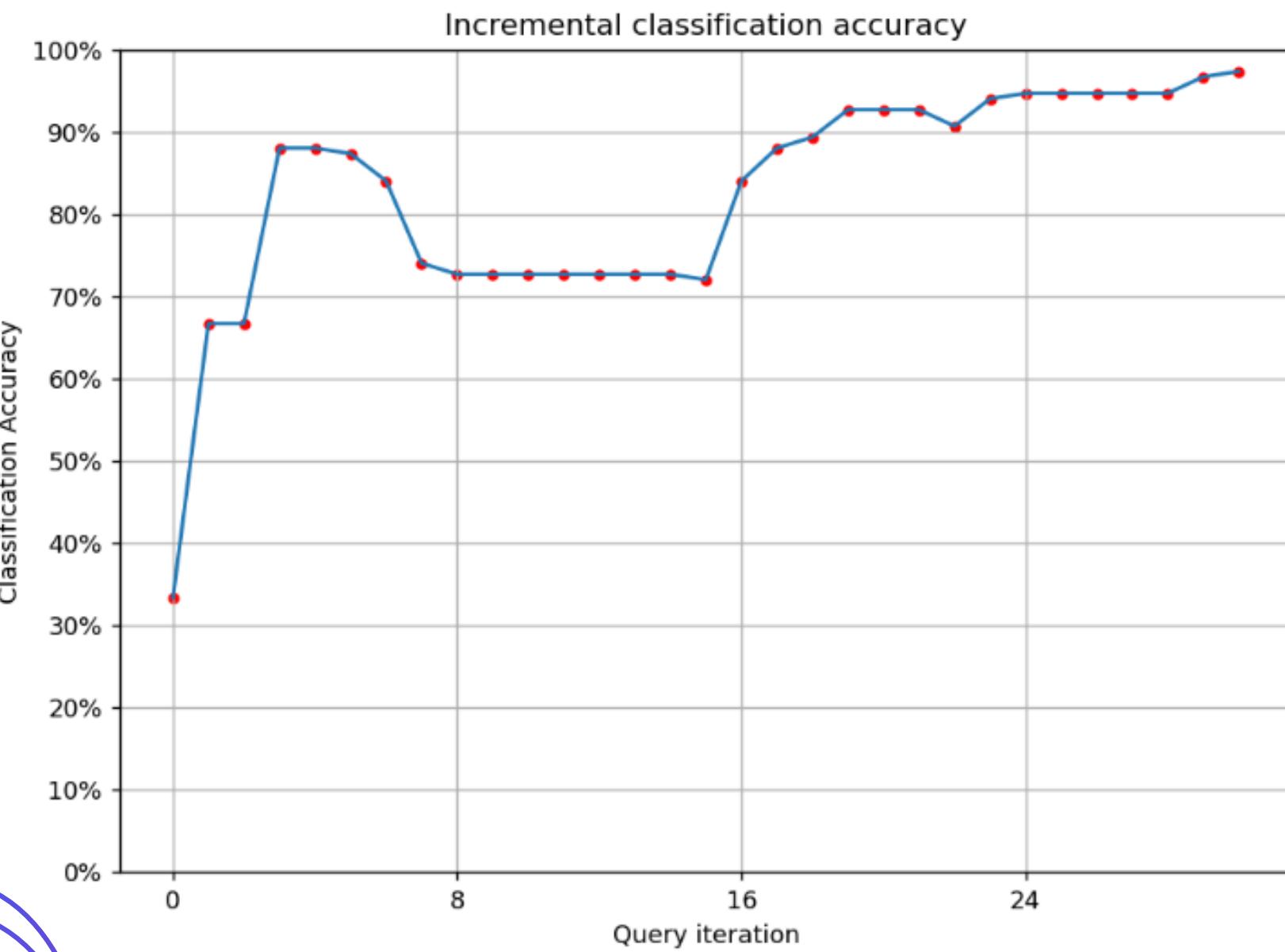
3

Random sampling

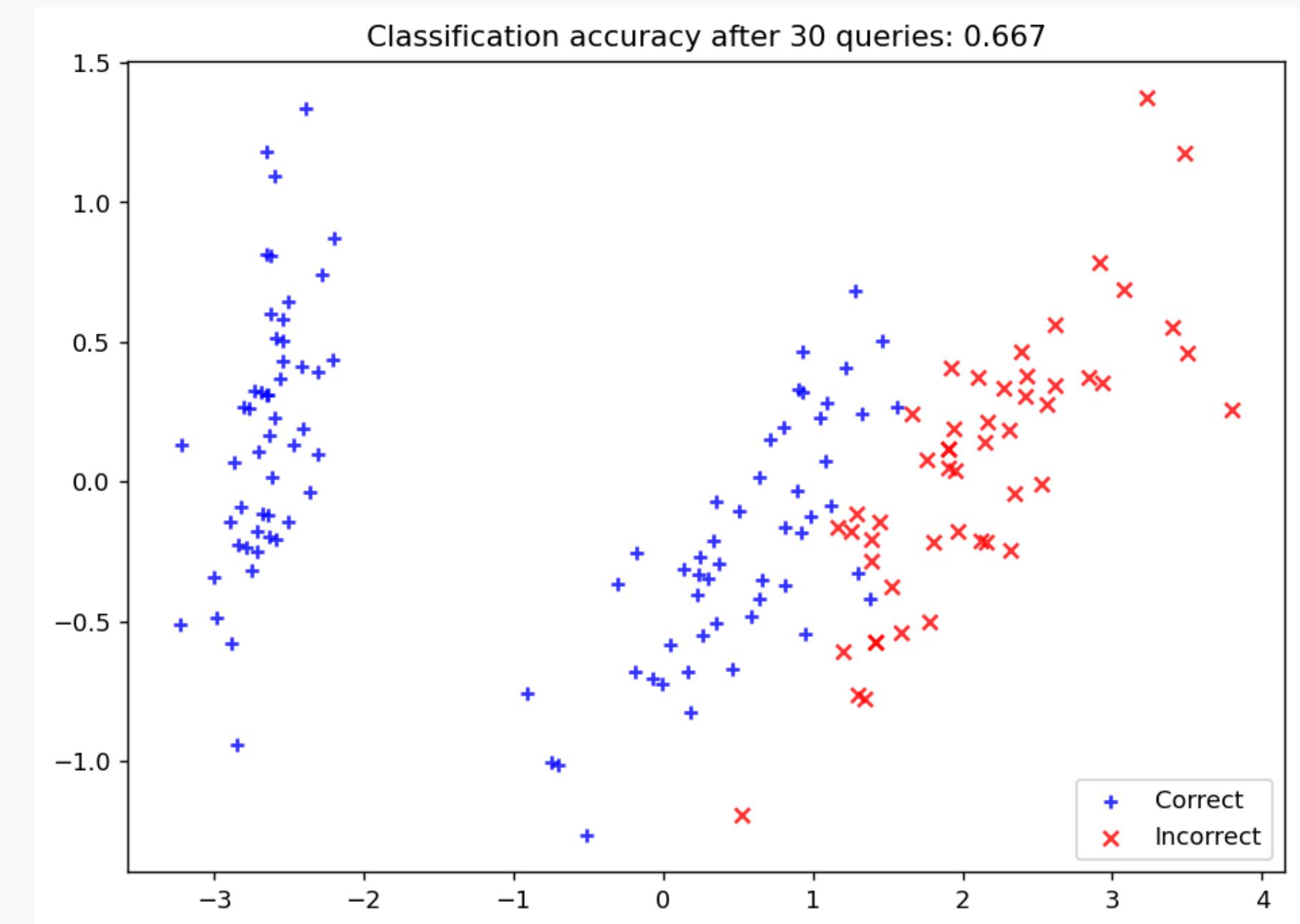
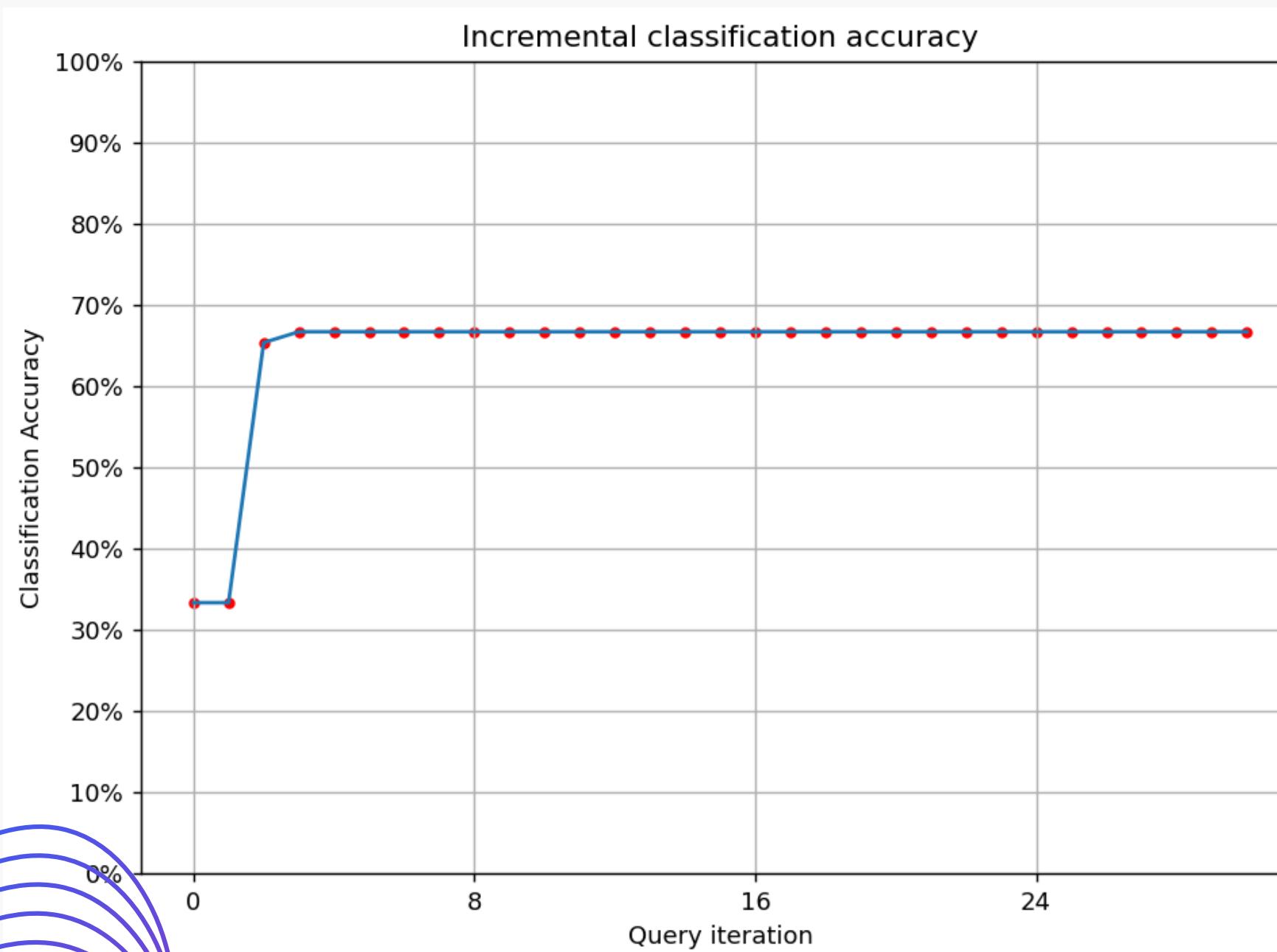


4

Entropy Sampling



5

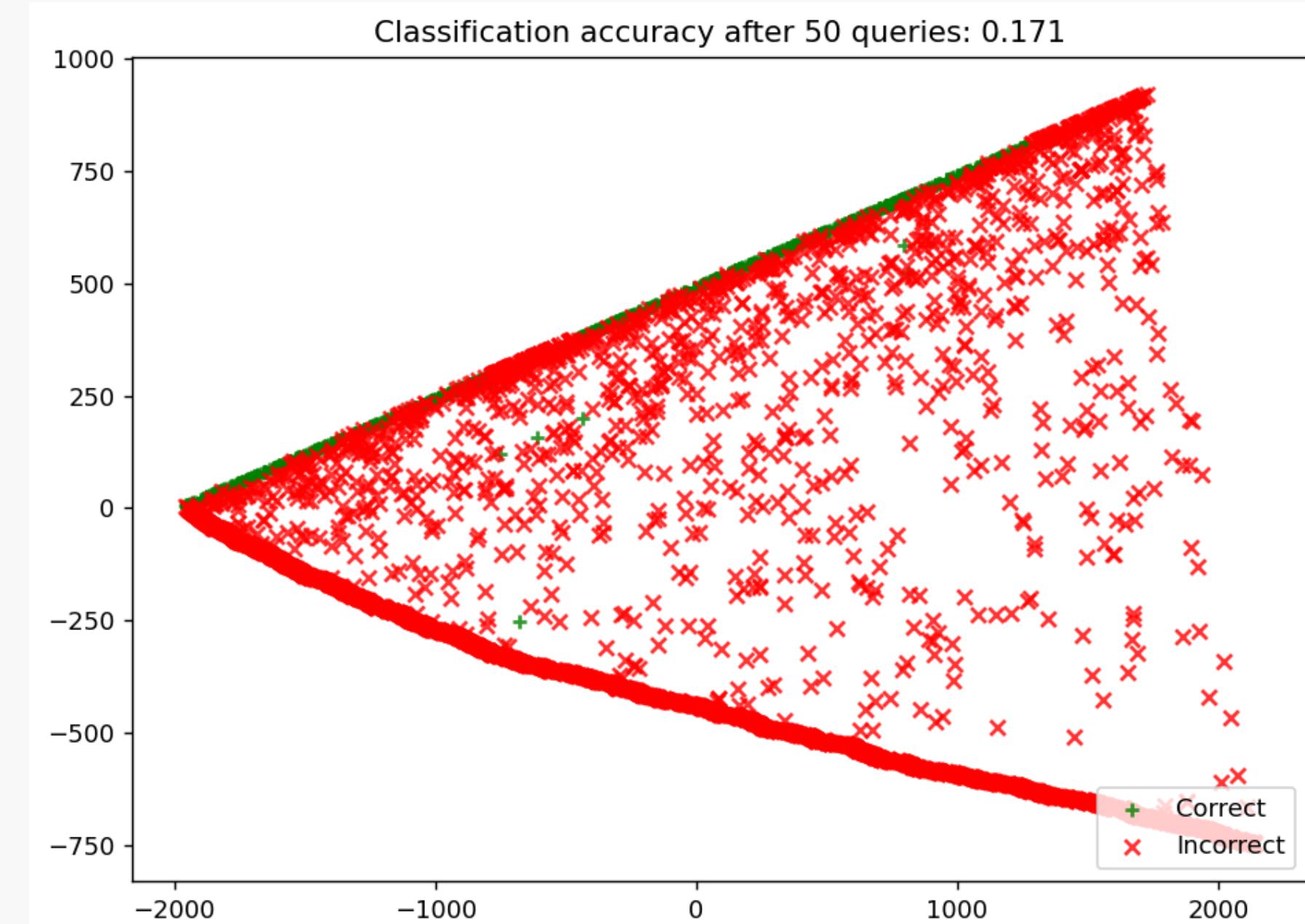
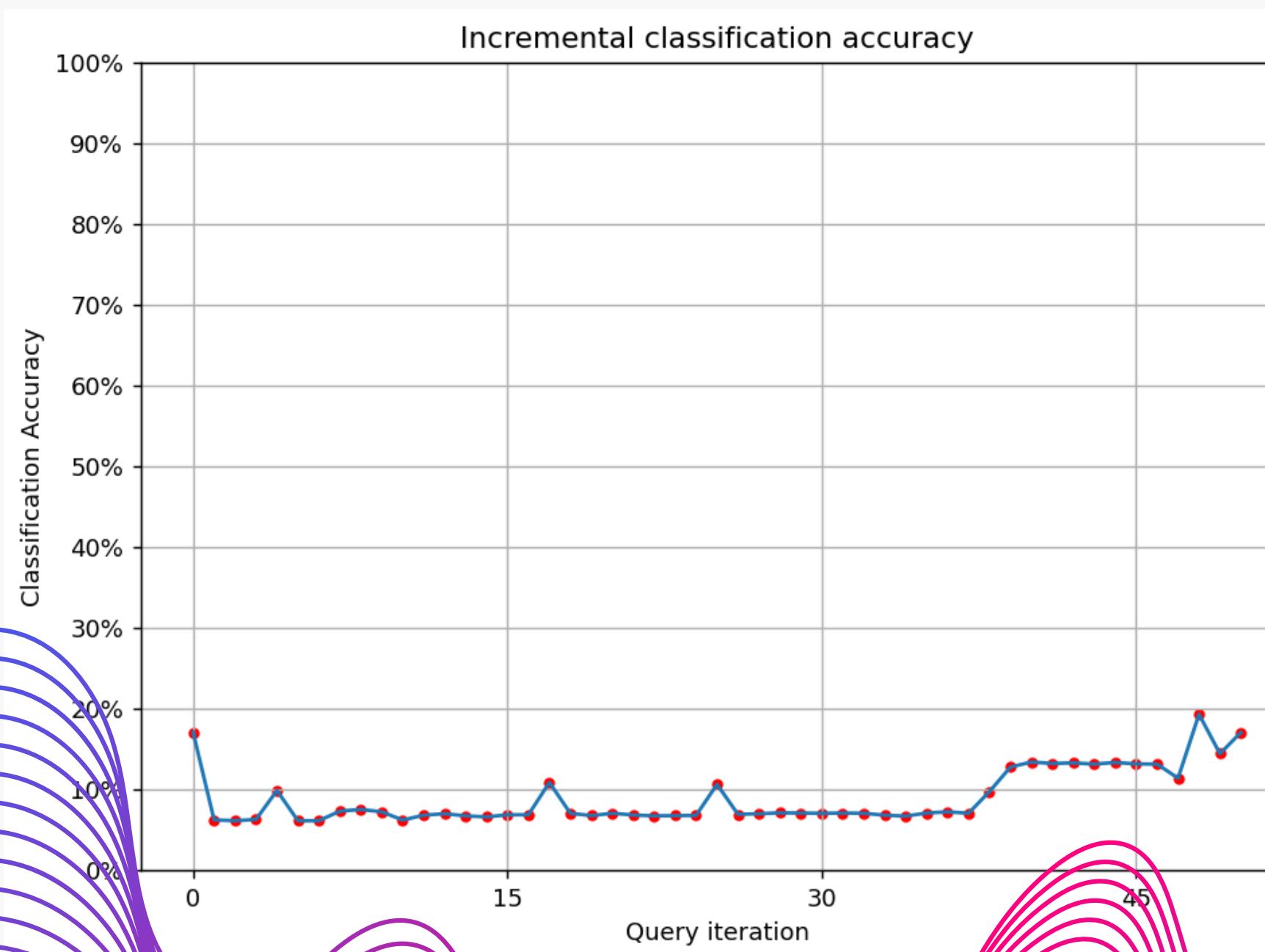
Uncertainty sampling

1

Corona_NLpdataset

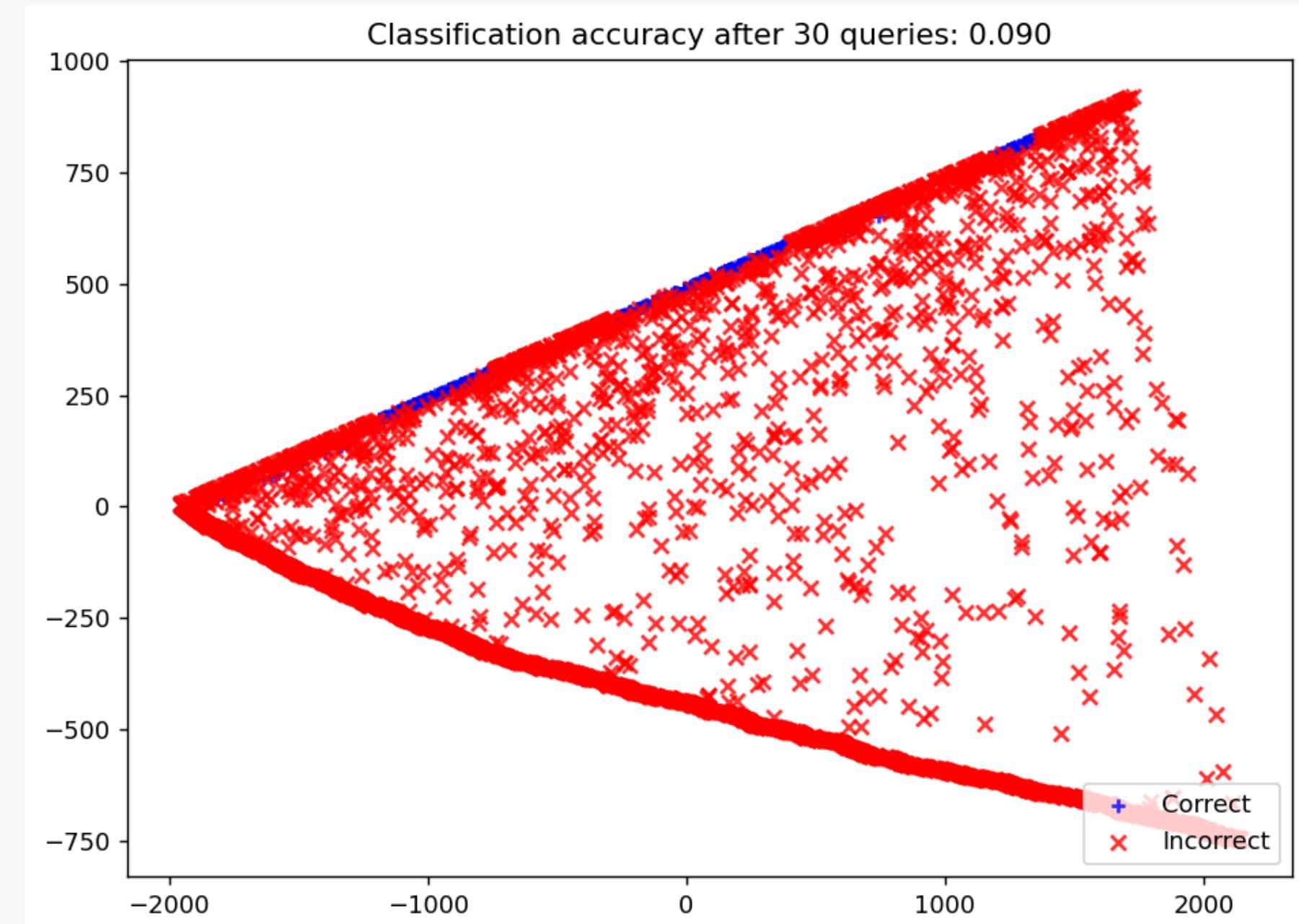
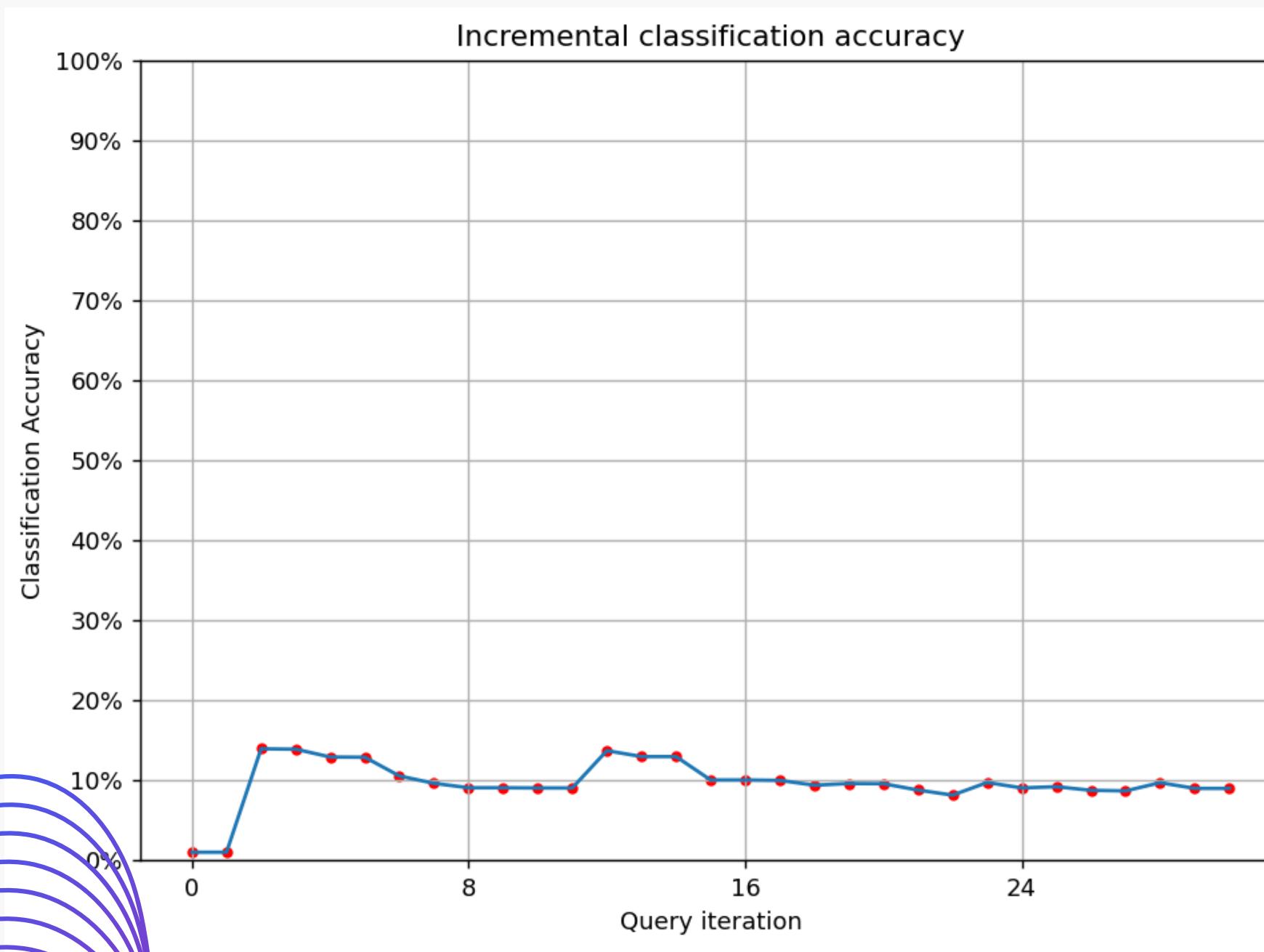
1

Margin Sampling



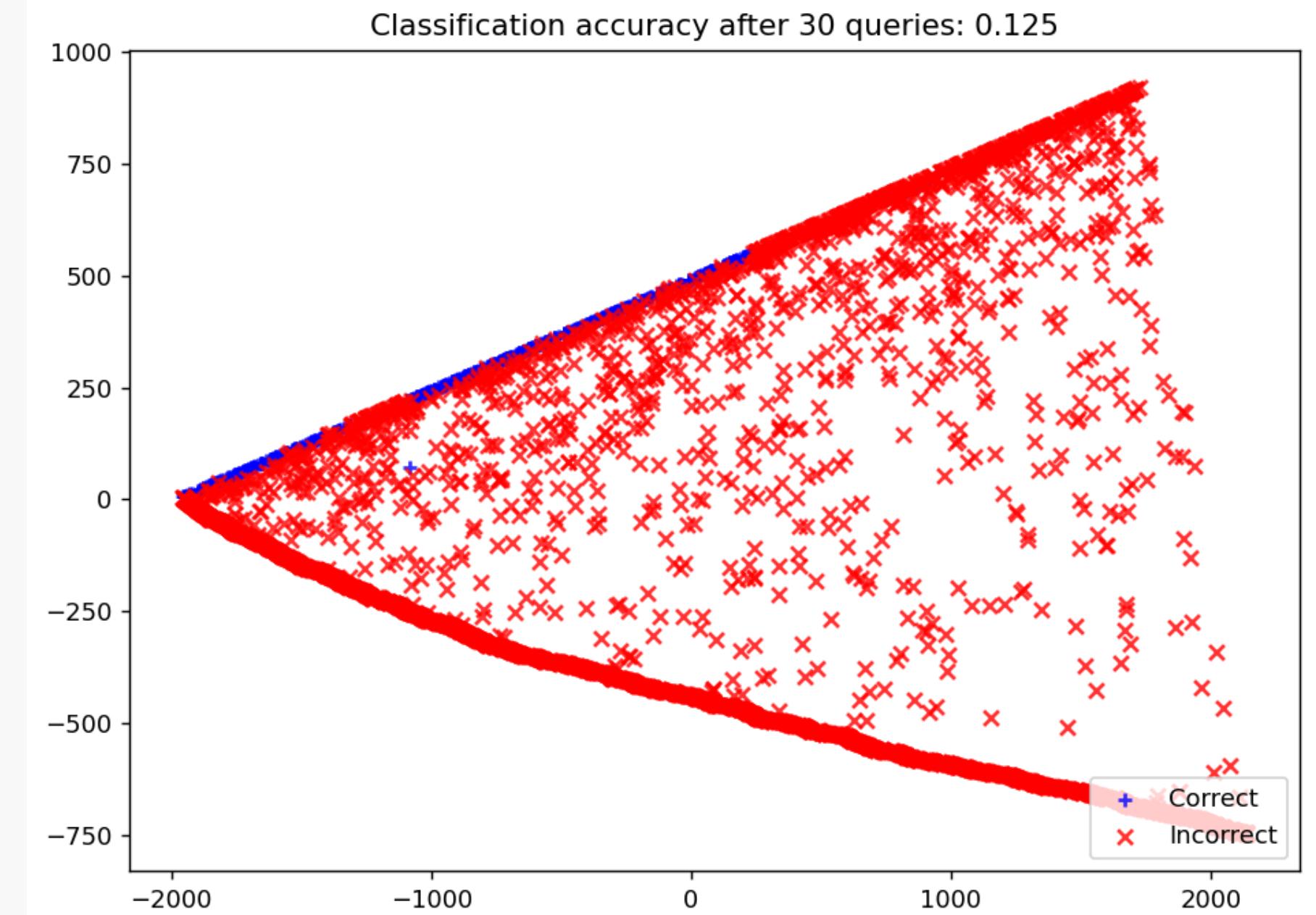
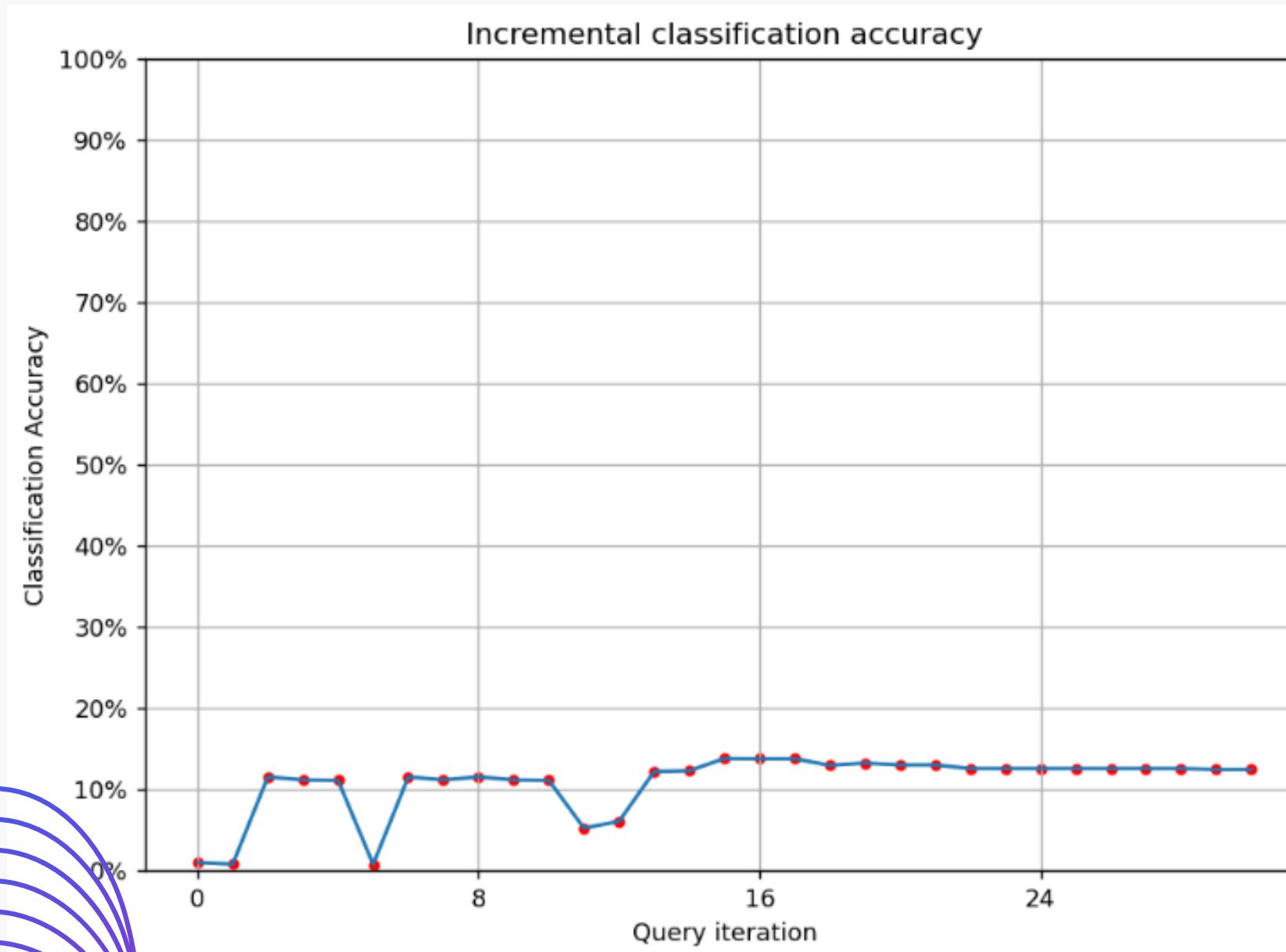
3

Random sampling

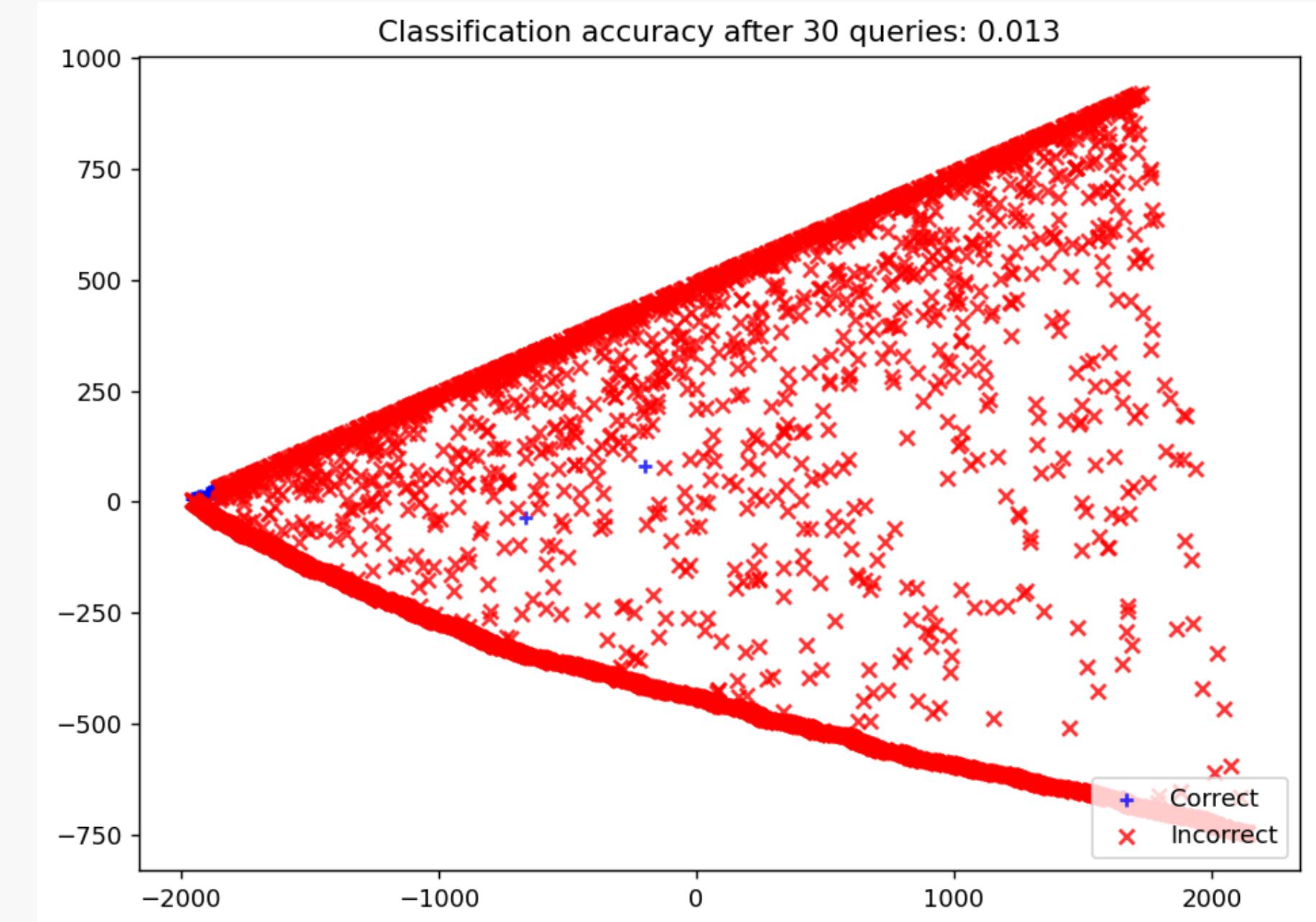
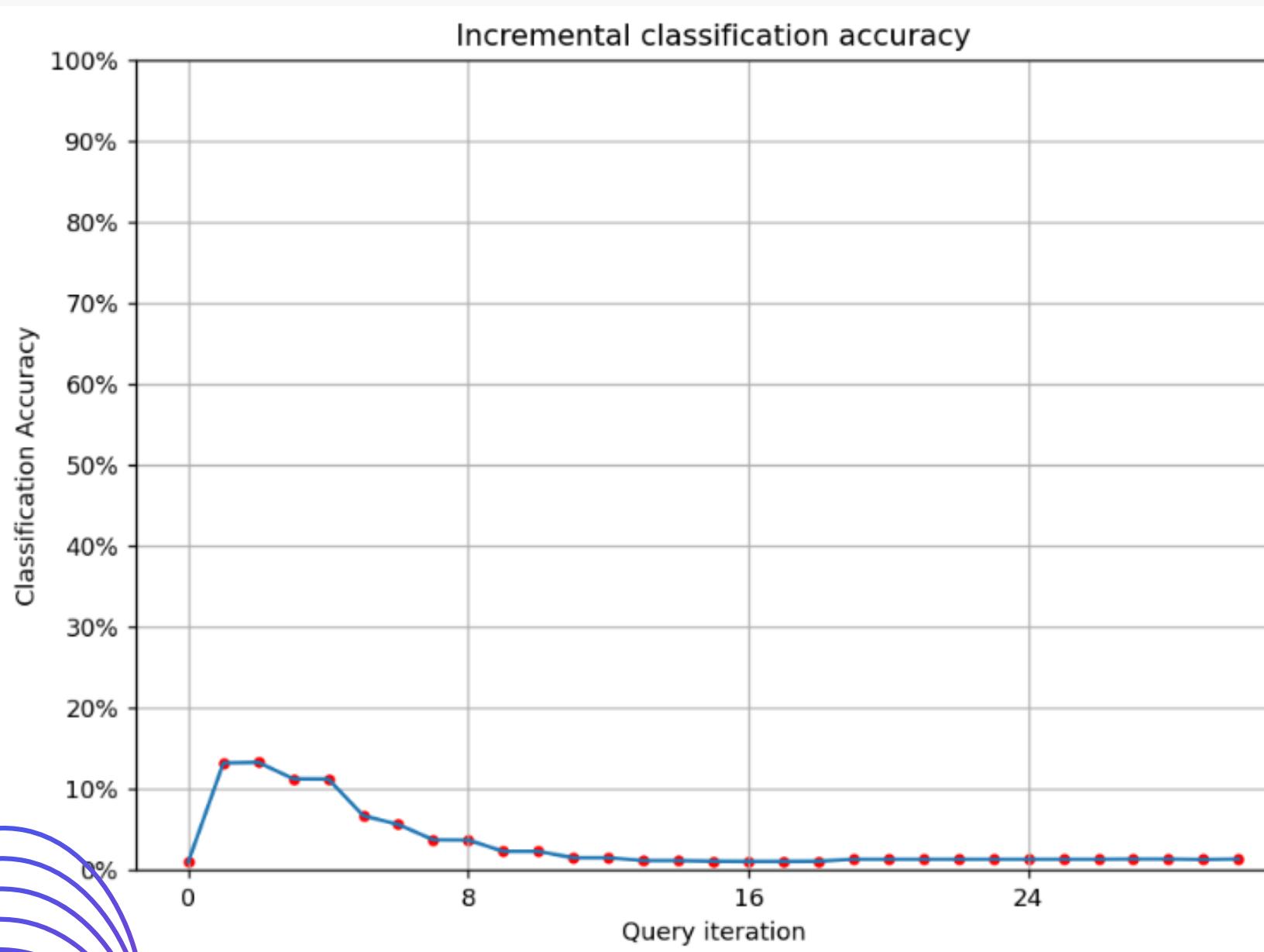


4

Entropy Sampling



5

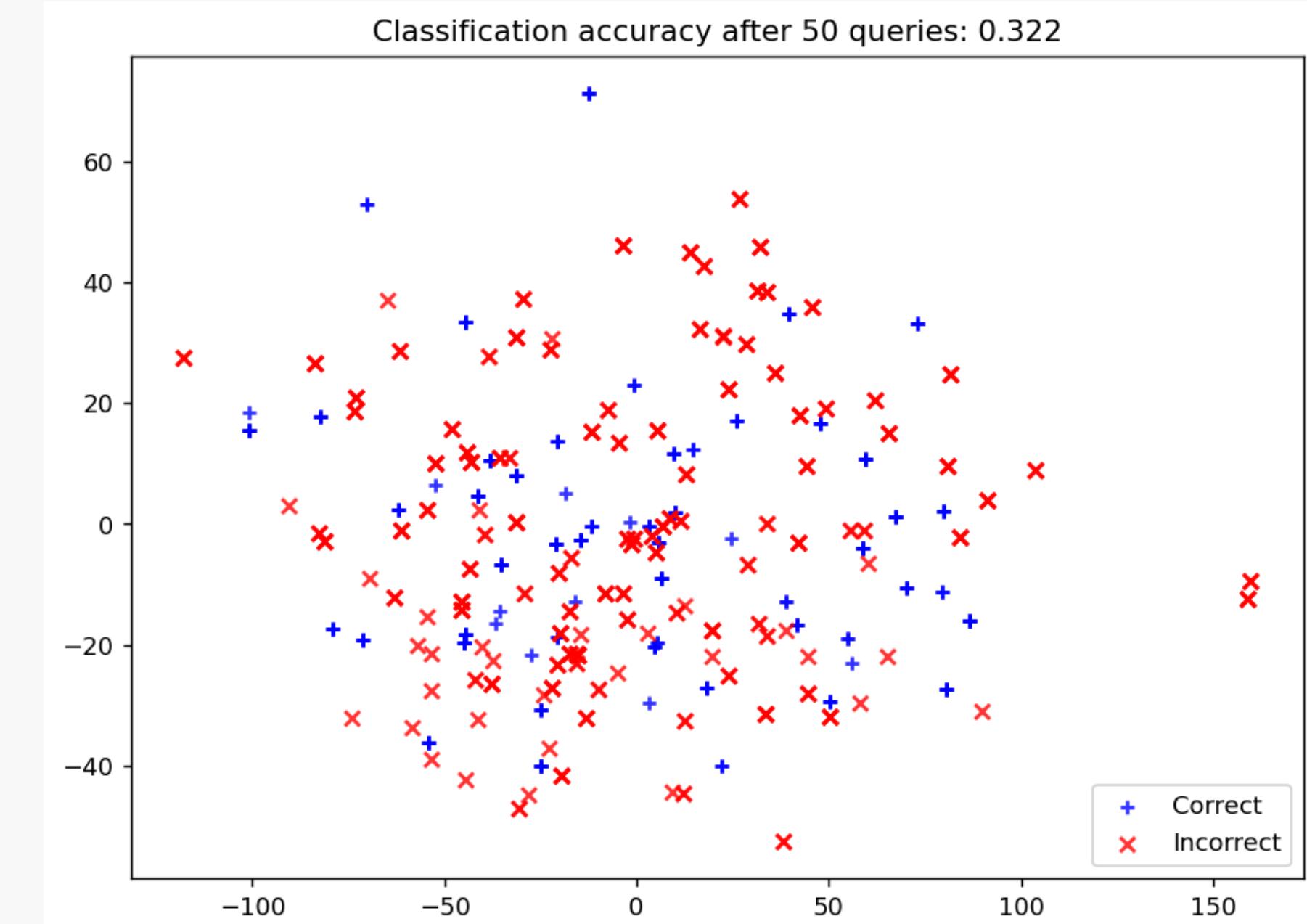
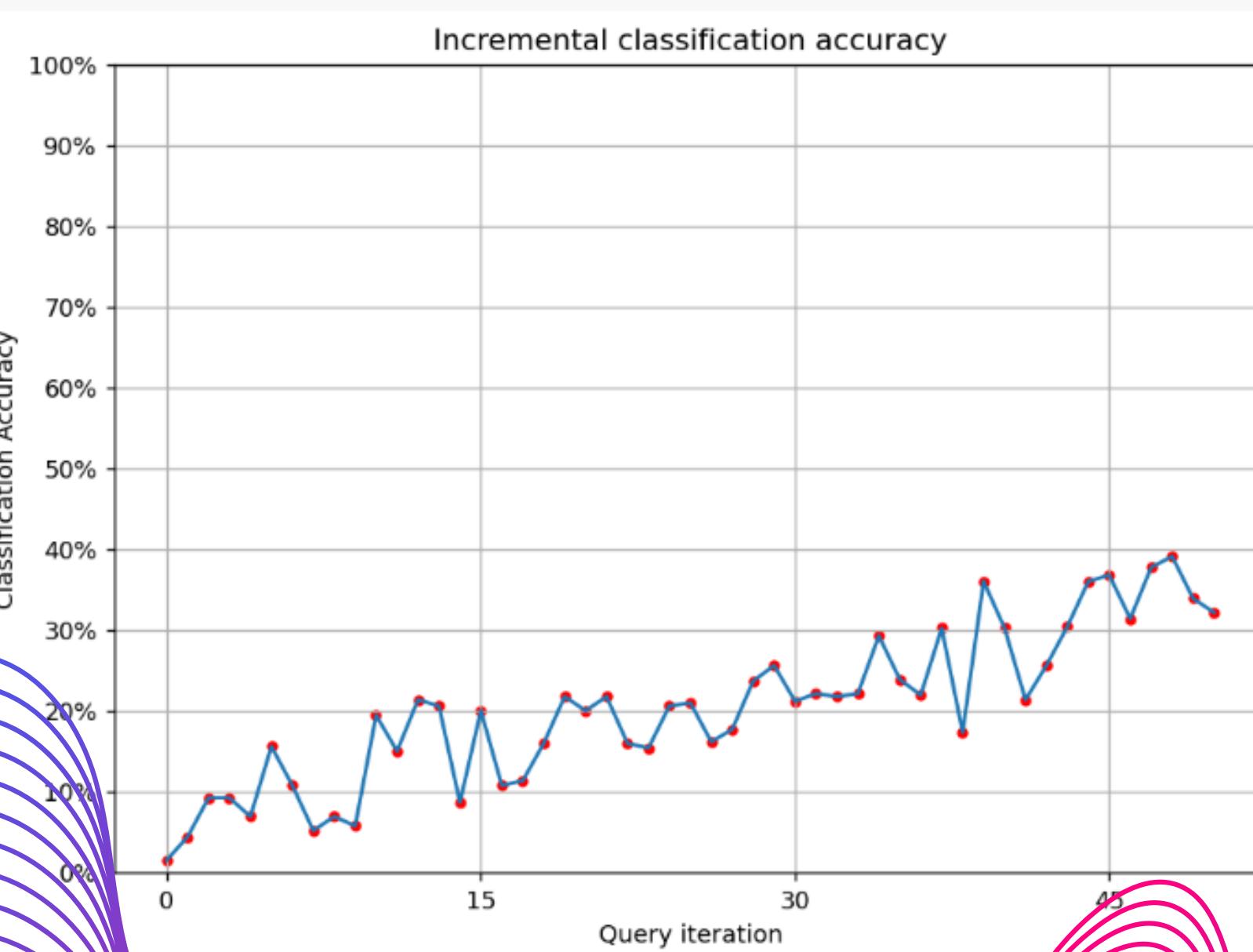
Uncertainty sampling

1

heart dataset

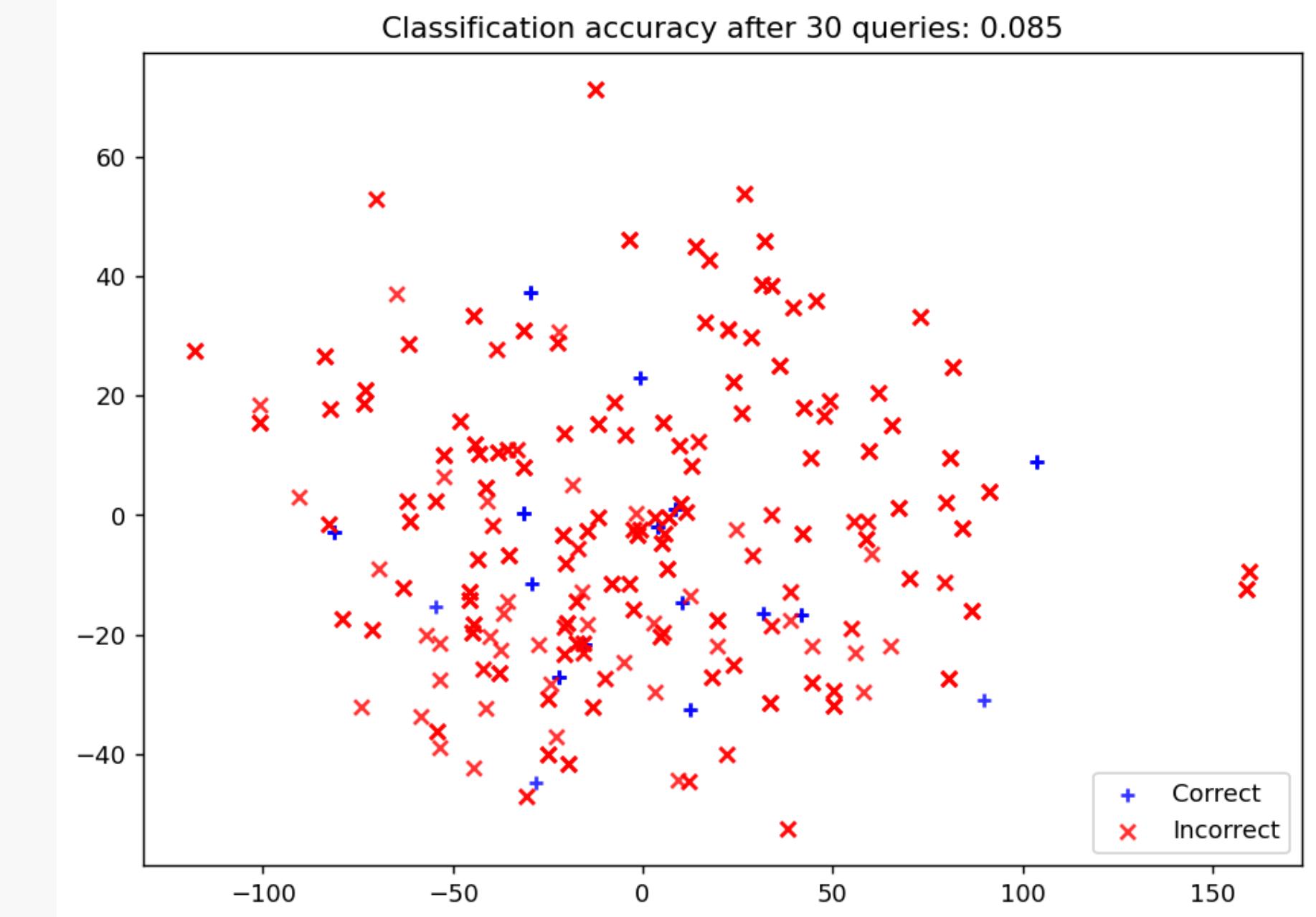
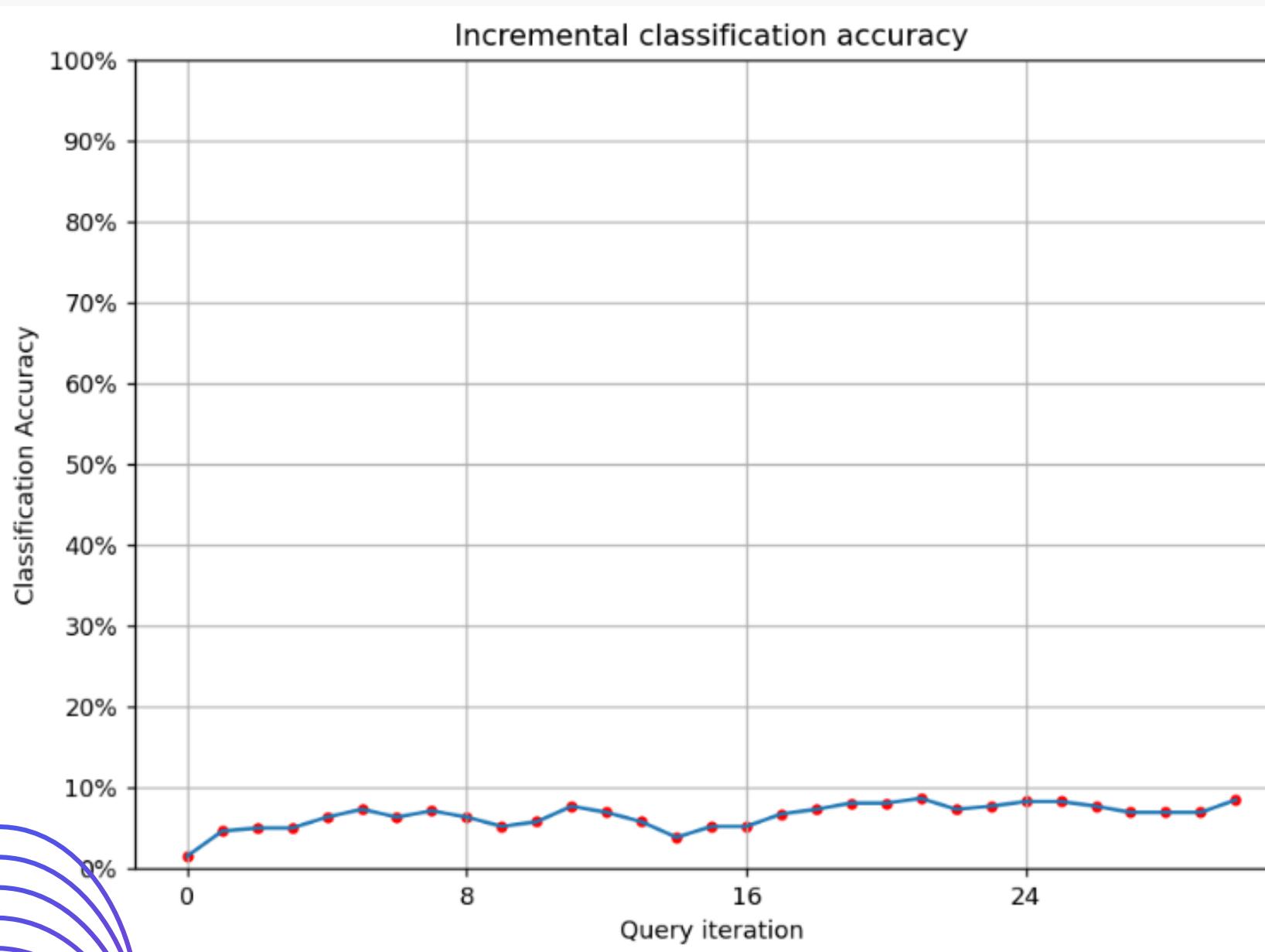
1

Margin Sampling

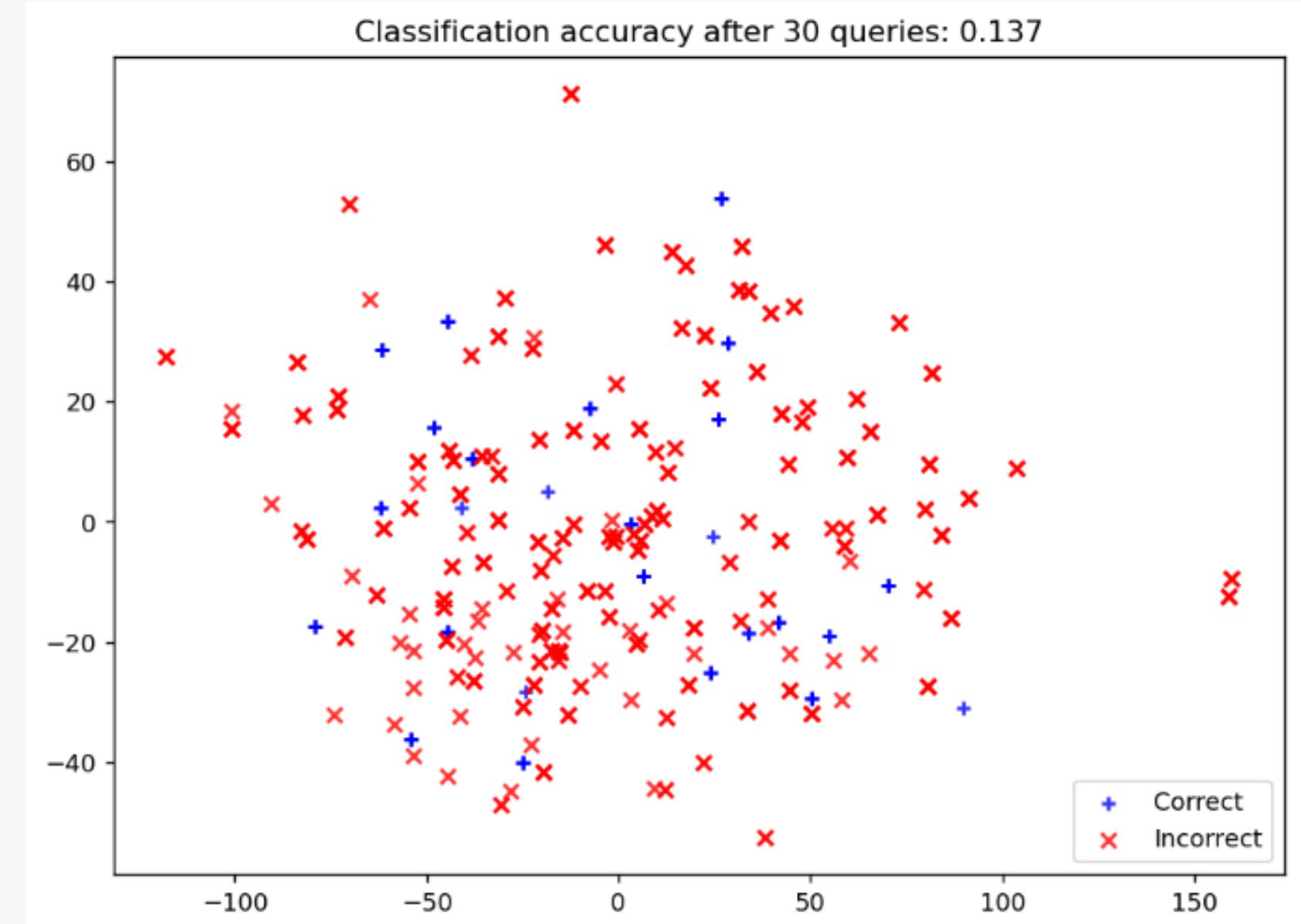
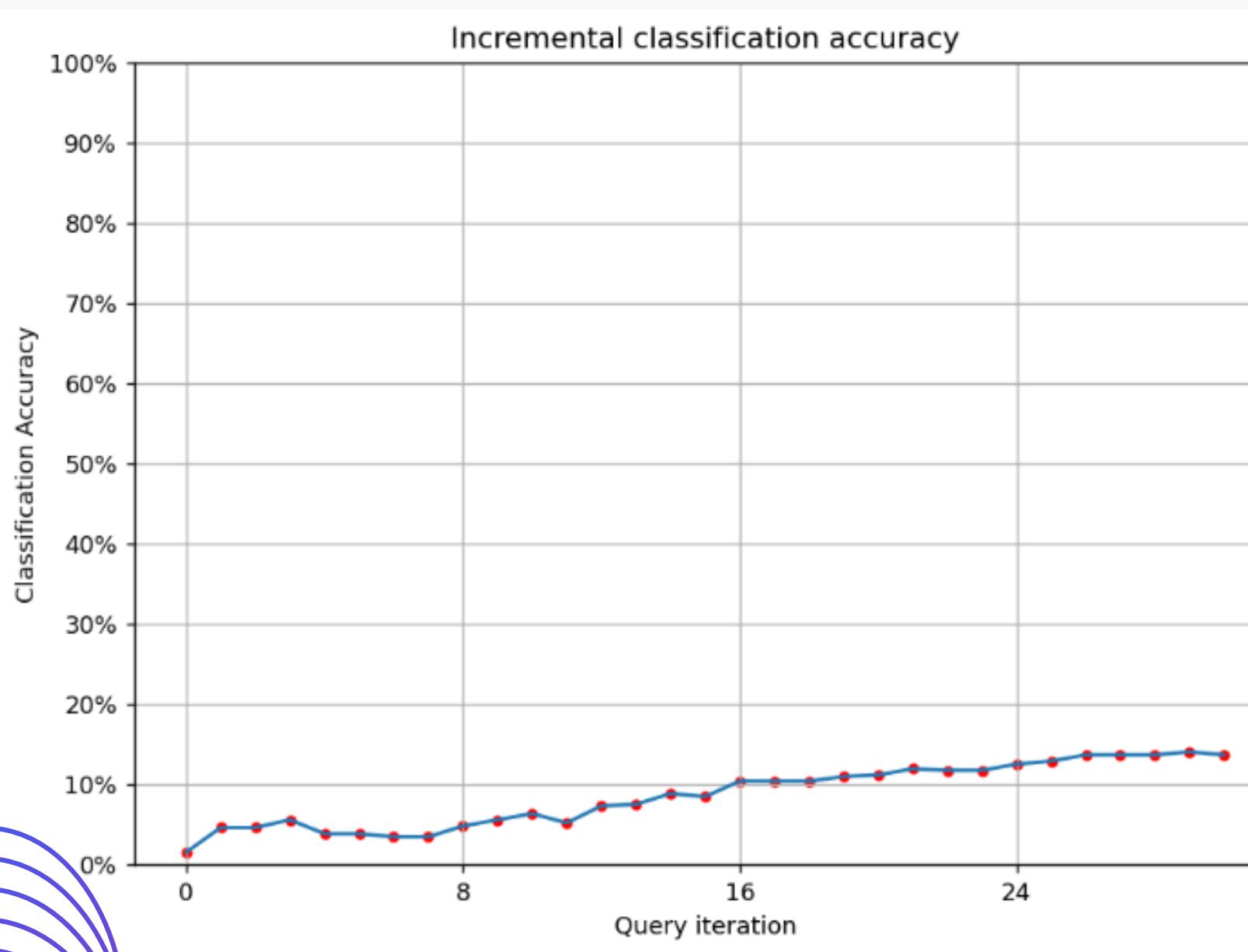


3

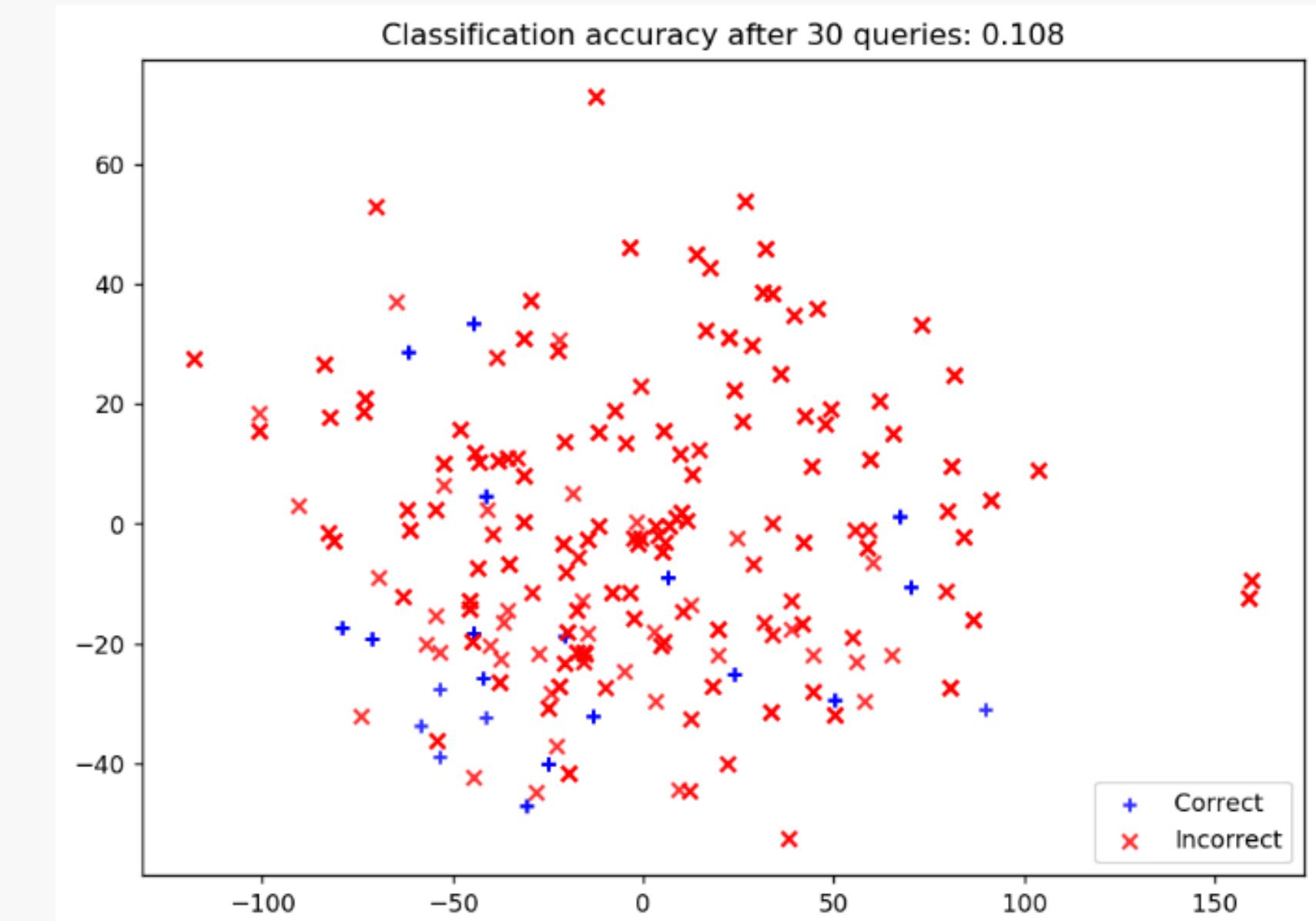
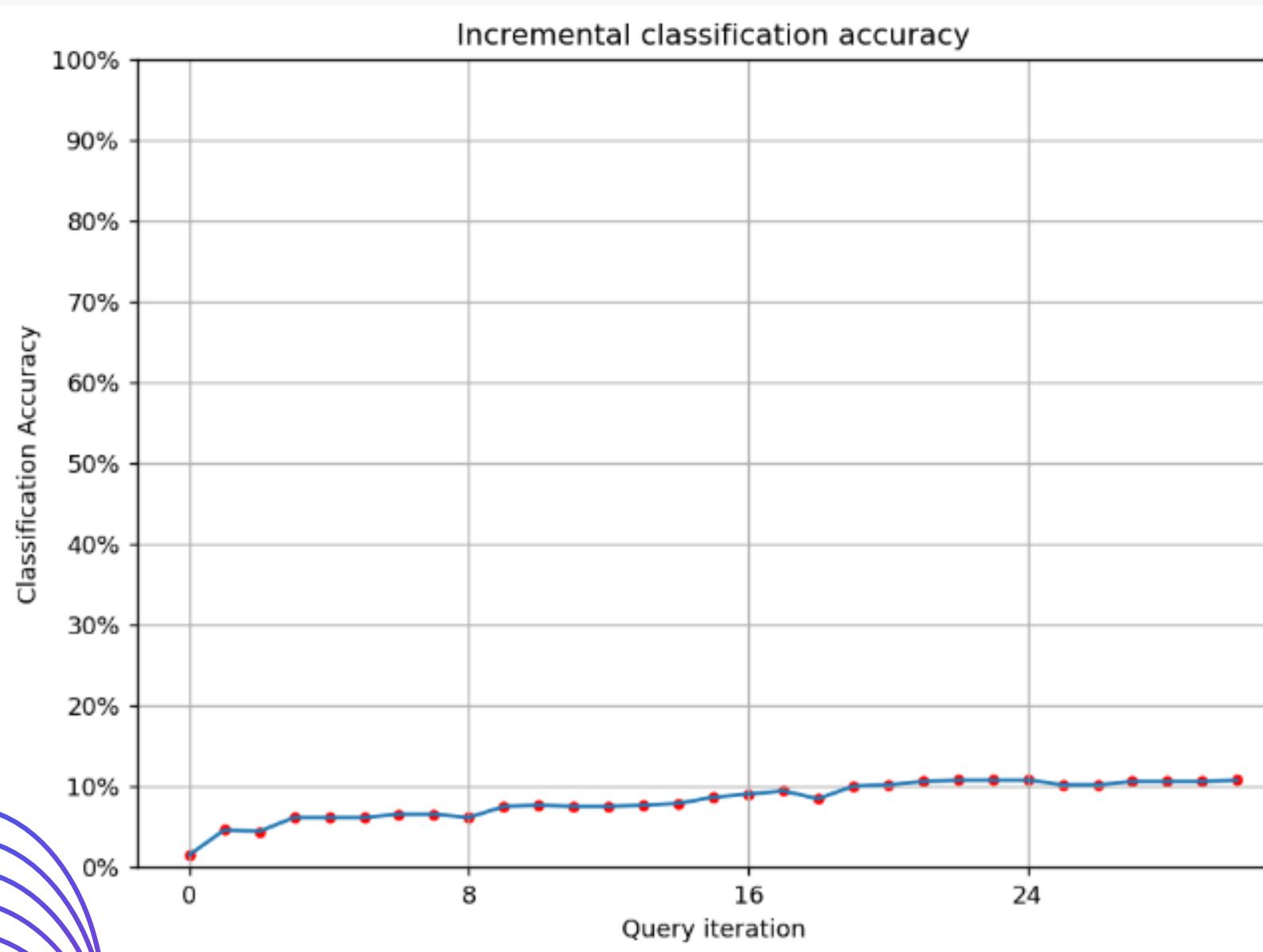
Random sampling



4

Entropy Sampling

5

Uncertainty sampling

**As we see that active learning has many strategies, these strategies
we can't say that one of them is the best because it dependent on
the data itself, so we should try many of them until find the best for
our data.**

*Thank
You*