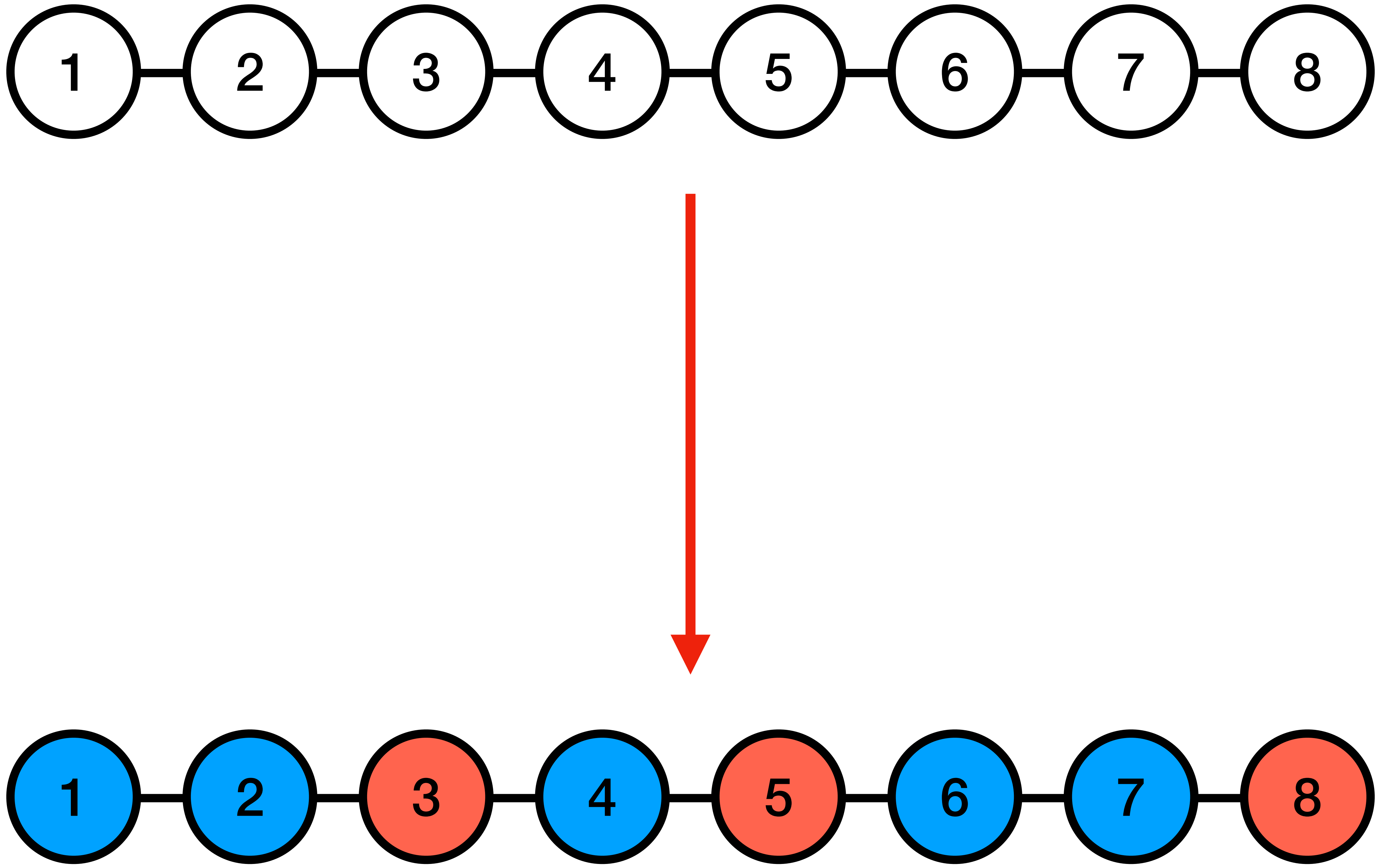
GALAXY: Graph-based Active Learning at the Extreme

Jifan Zhang, Julian Katz-Samuels, Robert Nowak
University of Wisconsin, Madison

Graph-based Active Learning

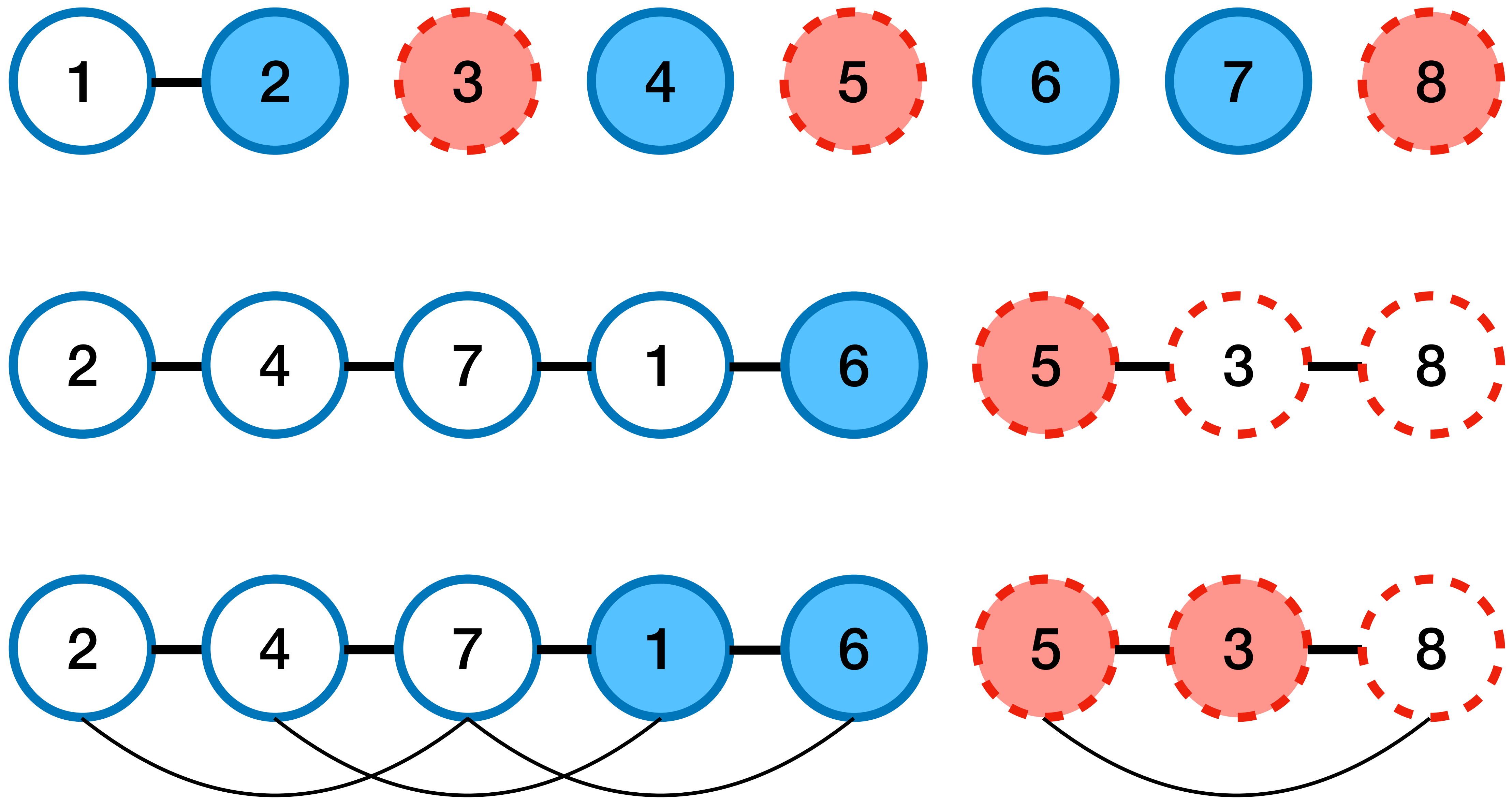
- Given a pool of unlabeled examples X, annotation cost is expensive
- Each example x_i is a node in graph $G = (X, E)$
- Each example has its corresponding label $f^\star(x_i)$
- **Objective: annotate as few examples as possible for a classifier to reach 100% accuracy.**



Insight from S^2

[Dasarathy, Nowak & Zhu, ICML 2015]

- Not all graphs are equally difficult (sample complexity) to learn.



Can we simultaneously learn
good classifiers and better graphs?

Leveraging a Deep Neural Network

Construct_Graph

Input: Pool X, number of classes K, neural network $f_\theta : \mathcal{X} \rightarrow \mathbf{\Delta}^{(K-1)}$

Initialize: Confidence for each $i \in [N]$:

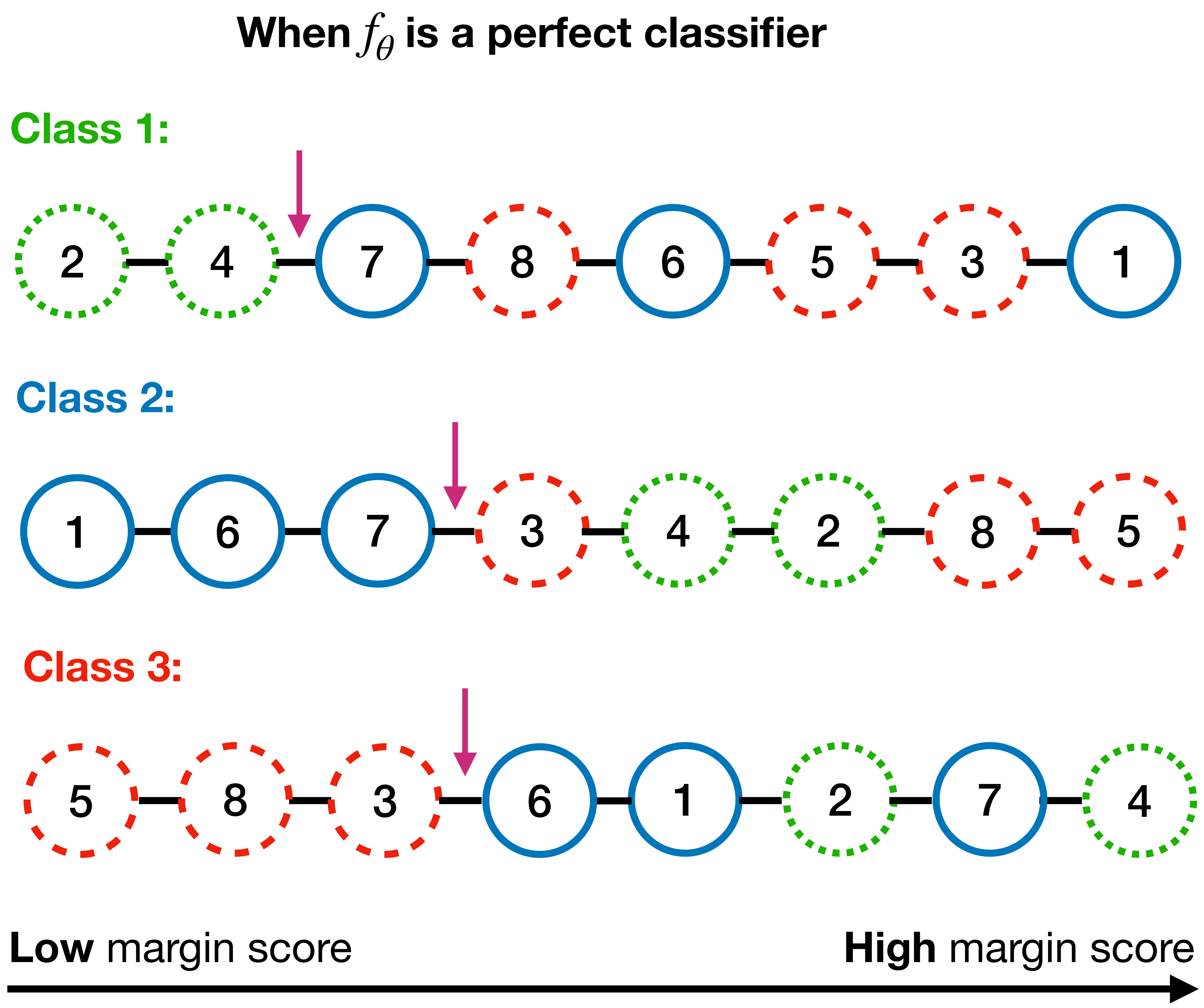
$$q_i \leftarrow \max_{k \in [K]} [f_\theta(x_i)]_k$$

For k = 1, …, K:

Sort examples by margin scores $\delta_i^{(k)} \leftarrow [f_\theta(x_i)]_k - q_i$ and break tie by q_i

Construct linear edge set $E^{(k)}$ connecting the sorted examples

Return: Graphs $\{G^{(k)} = (X, E^{(k)})\}_{k=1}^{K}$



Leveraging a Deep Neural Network

Construct_Graph

Input: Pool X, number of classes K, neural network

$$f_\theta : \mathcal{X} \to \mathbf{\Delta}^{(K-1)}$$

Initialize: Confidence for each $i \in [N]$:

$$q_i \leftarrow \max_{k \in [K]} [f_\theta(x_i)]_k$$

For k = 1, …, K:

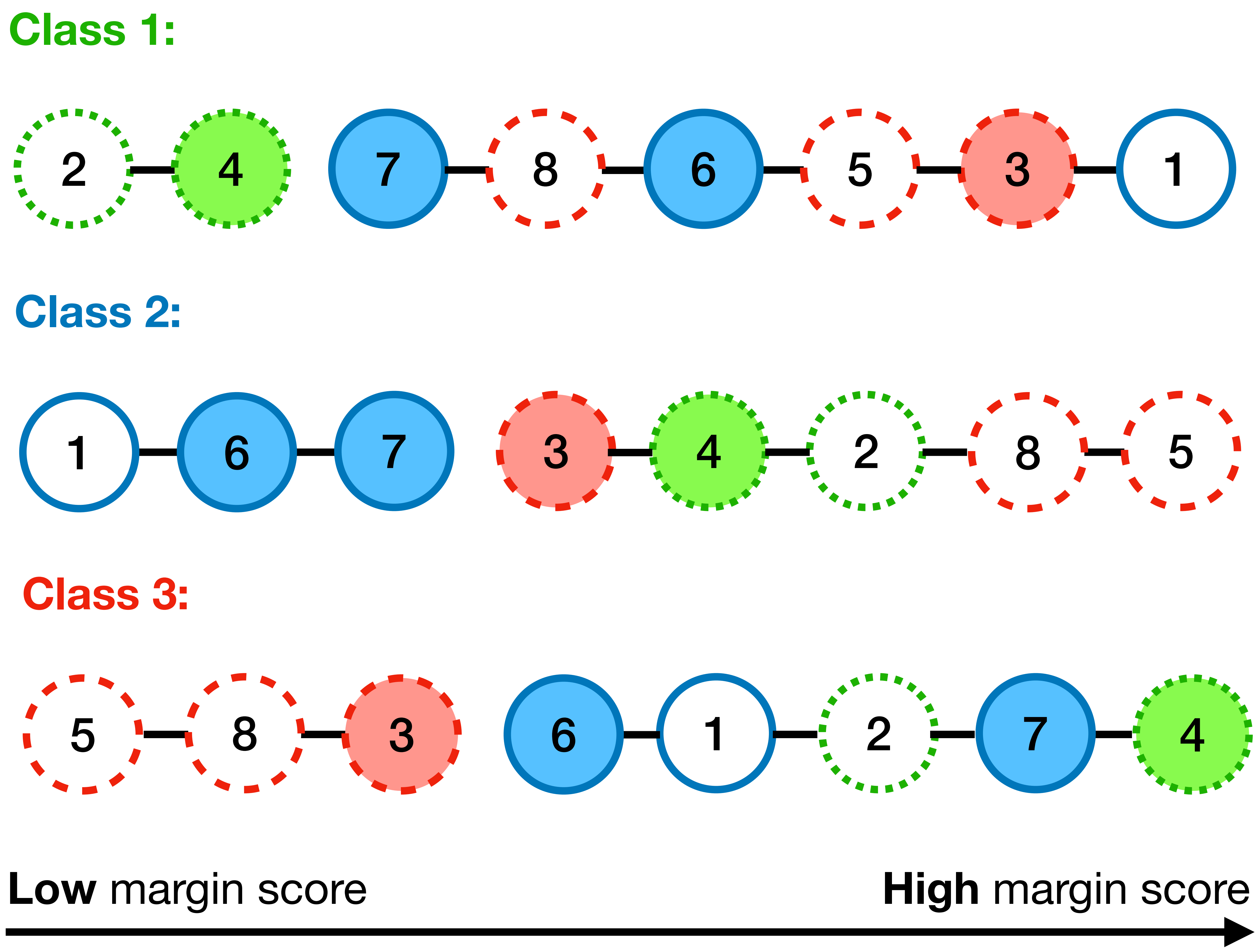
Sort examples by margin scores

$\delta_i^{(k)} \leftarrow [f_\theta(x_i)]_k - q_i$ and break tie by q_i

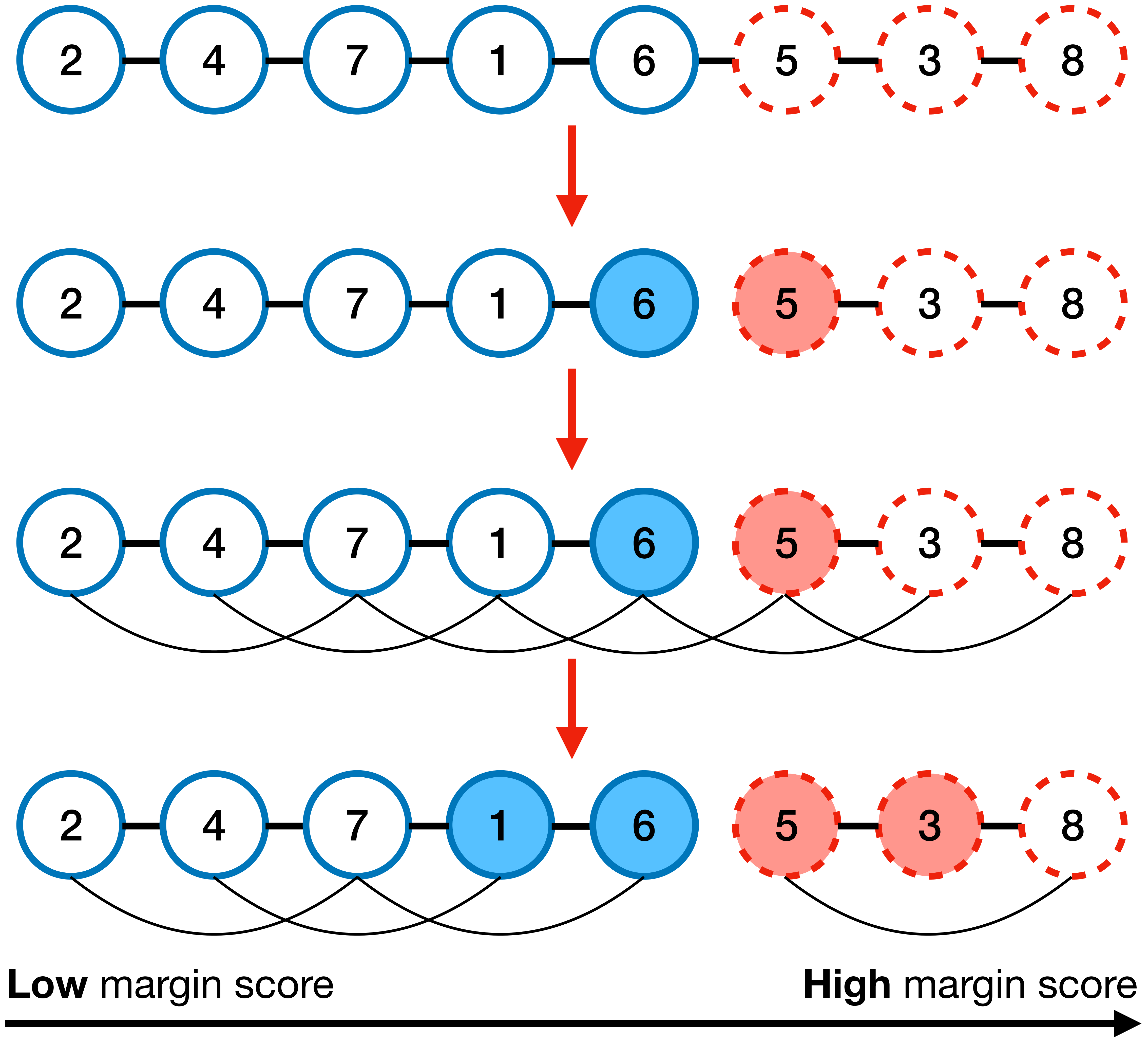
Construct linear edge set $E^{(k)}$ connecting the sorted examples

Return: Graphs $\{G^{(k)} = (X, E^{(k)})\}_{k=1}^K$

Idea: Bisect to find one-vs-all cuts



GALAXY in Separable Case
= Bisection + Querying Around Uncertainty Boundary



Extreme Class Imbalance

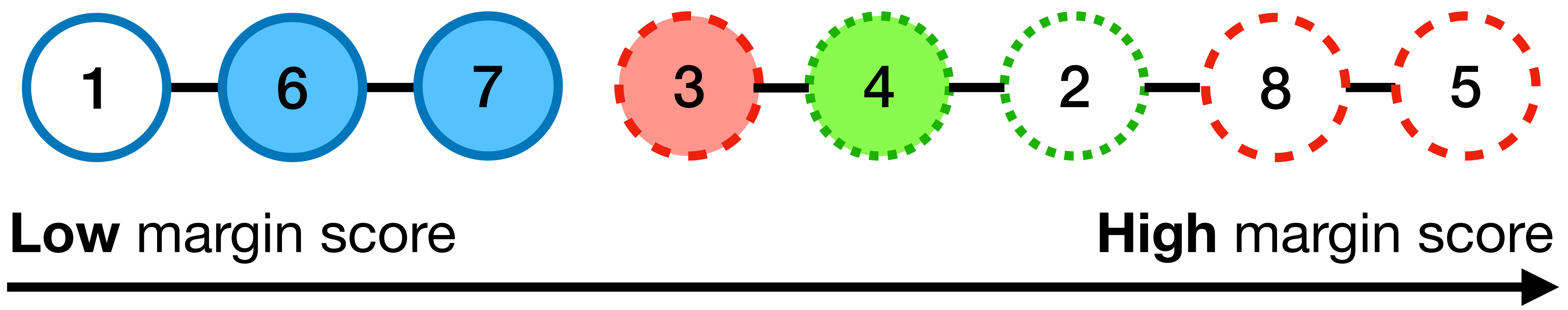
- Define number of examples in each class: $N_k = |\{x \in X | f^\star(x) = k\}|$
- Extreme class imbalance

$$\frac{N_k}{N_K} \leq \epsilon,\ k = 1,\ldots,K-1$$

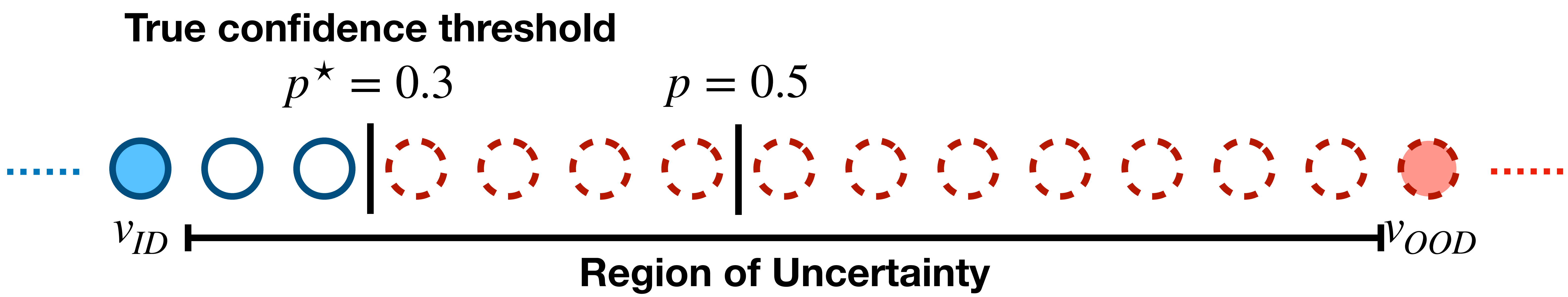
I.e., one class has significantly more examples than any other class.

Key Ingredients

- Annotate Uncertain Examples



- [Theorem] Annotate Class-diverse (class-balanced) Examples: lower bound on the balancedness of collected examples



- [Theorem] Uncertainty sampling in certain cases could collect **zero** minority labels on expectation.

Experiments: Extremely Unbalanced Datasets

- We modify CIFAR-10, CIFAR-100 and SVHN datasets.
  - If K classes in total, then class $1,\ldots,K-1$ are the original classes while class K includes the rest of the original dataset.
  - For example, for CIFAR-100 with 3 classes, we have classes 1 vs 2 vs 3~100.

Experiments: Extremely Unbalanced Datasets

NAME	# CLASSES	N_K	$\sum_{k=1}^{K-1} N_k$	ϵ
CIFAR-10	2	45000	5000	.1111
CIFAR-10	3	40000	10000	.1250
CIFAR-100	2	49500	500	.0101
CIFAR-100	3	49000	1000	.0102
CIFAR-100	10	40500	9500	.0123
SVHN	2	68309	4948	.0724
SVHN	3	54448	18809	.2546
PATHMNIST	2	80595	9401	.1166

Table 1. Dataset details for each extremely unbalanced scenario. N_K denotes the number of images in the out-of-distribution class while $\sum_{k=1}^{K-1} N_k$ is the total number of images in all in-distribution classes. ϵ is the class imbalance factor defined in Section 3.

Baselines

Uncertainty-based Algorithms

Confidence Sampling [Settles, 2009]

Cluster Margin [Citovsky et al., NeurIPS 2021]

Badge [Ash et al., ICLR 2020]

BAIT [Ash et al., NeurIPS 2021]

Class-diverse Algorithms

Most Likely Positive [Jiang et al., NeurIPS 2018; Warmuth et al., NIPS 2001; 2003]

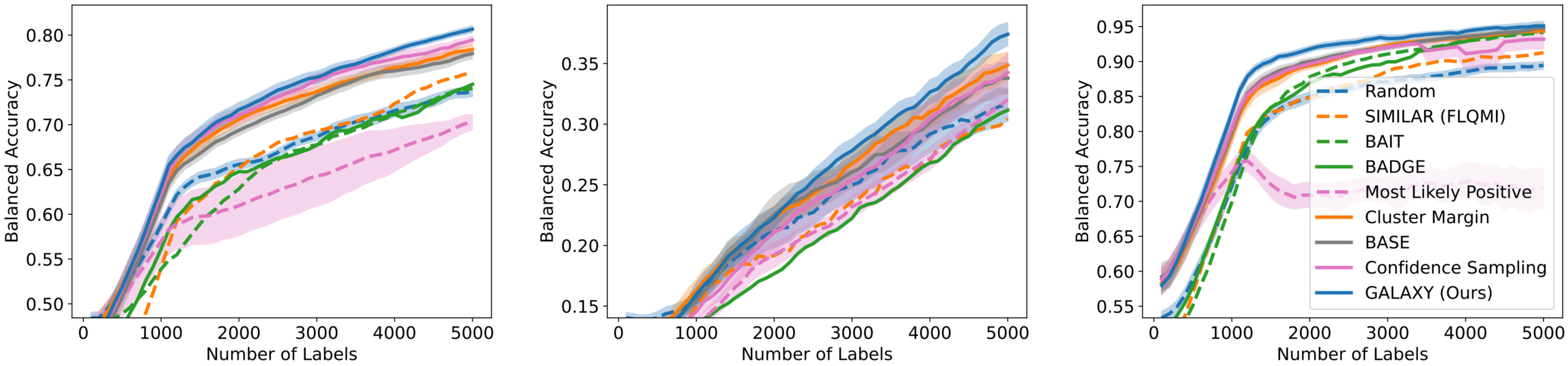
SIMILAR [Kothawade et al., NeurIPS 2021]

Uncertainty & Class-diverse Algorithms

BASE [Emam, 2021]

GALAXY

Results: our algorithm outperforms in all settings.



(a) ACC_{bal}, CIFAR-10, 3 classes (b) ACC_{bal}, CIFAR-100, 10 classes (c) ACC_{bal}, SVHN, 2 classes

Figure 4. Performance of `GALAXY` against baselines on selected settings. Legend shown in (c) is shared across all three plots.

Conclusion

- GALAXY vs S^2: Utilize deep neural network training to actively improve graphs.
- GALAXY vs other deep active learning
    - Finds the right uncertainty threshold and queries ***uncertain*** examples
    - Collects more ***balanced*** (class-diverse) labelled set
- Future work: Currently a batch is sequentially labelled. Can we parallelize the process to allow multiple simultaneous annotators?