Outlier Injection in K-means Clustering

Poison Cluster Team

Budapest, 2025.03.25

Problem Statement

- Data poisoning attacks: inserting false data to manipulate clustering results
- Challenges in detecting poisoned data
- Need for robust clustering algorithms

Selected Clustering Algorithm

- K-Means Clustering
  - Partition-based clustering
  - Sensitive to outliers and initialization

Dataset

- Iris dataset
- Well-known dataset for classification/clustering
- Features:
  - Sepal length, sepal width, petal length, petal width
- Three natural clusters (species)

Attack Strategies

- Data Poisoning Attacks
  - Injecting false data to alter cluster formation
  - Creating adversarial examples
- Outlier Injection: Adding extreme values to distort clusters

Evaluate Attack

- Evaluate cluster differences
- Investigate detection techniques
- Assess the severity and risks of the attack

Detection Approach

- Clustering-Based Detection
- Identify anomalies with K-Means
- Compare clustering results before and after poisoning

ELTE EÖTVÖS LORÁND UNIVERSITY

Thank you for your attention!