ADVERSARIAL MACHINE LEARNING

TAXONOMY OF ADVERSARIAL ATTACKS

8

- Attack Position Based
    - Poisoning
        - Perturbation of the training data
    - Evasion
        - Crafting adversarial examples at the testing phase

RAW DATA

FEATURE REPRESENTATION

CLASSIFICATION

Poisoning

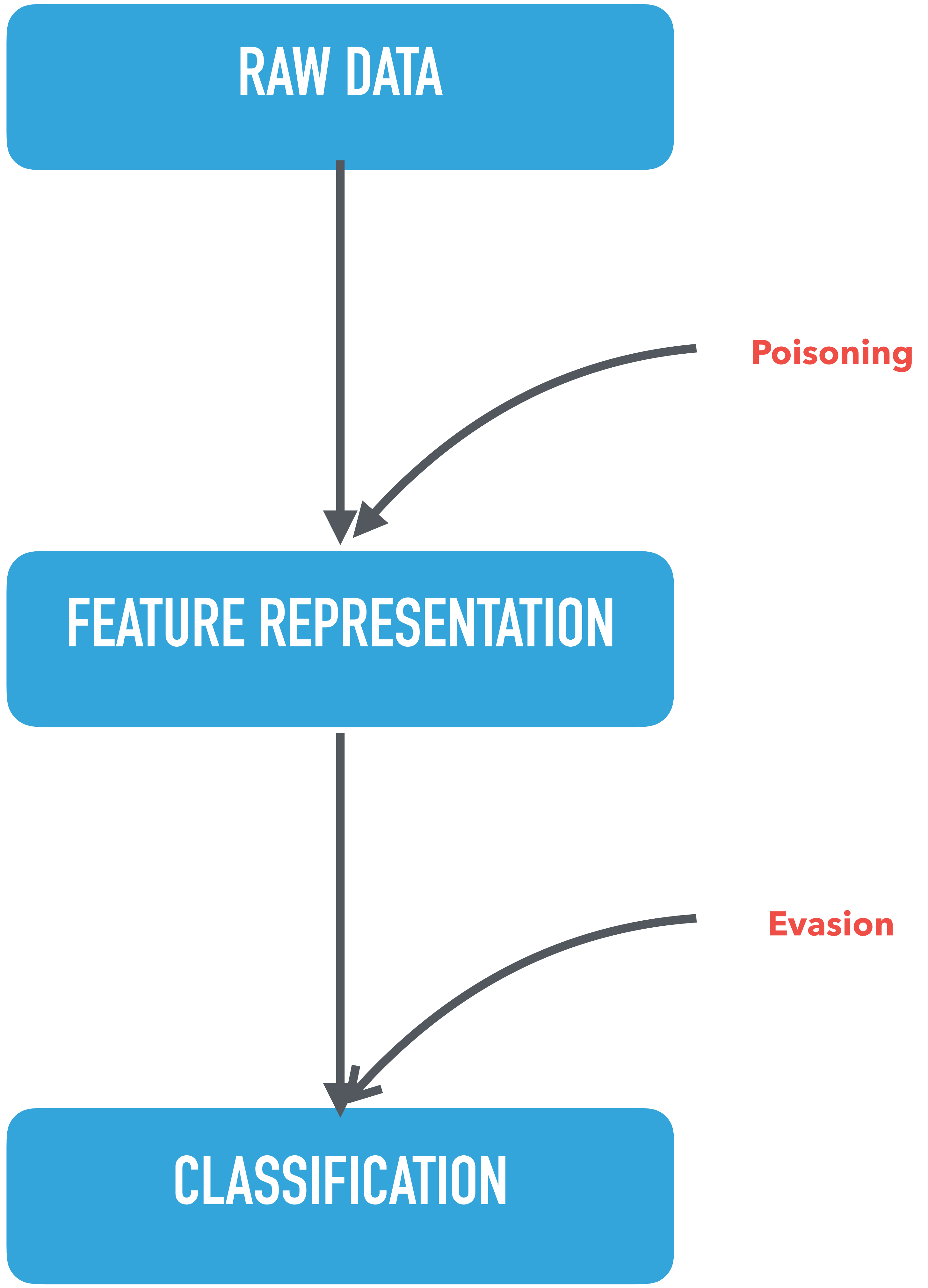
Evasion

TAXONOMY OF ADVERSARIAL ATTACKS

- Based on the intention of the adversary
  - Targeted
    - Multi class (Goal of adversary to have an adversarial example getting classified into a particular class)
      - e.g. Classify stop sign as speed up
    - Fooling anomaly detection (special case - binary classification)
  - Non-Targeted
    - Goal of adversary is to just break the classifier

TAXONOMY OF ADVERSARIAL ATTACKS

- Attack Position Based
  - Poisoning
    - Perturbation of the training data
  - Evasion
    - Crafting adversarial examples at the testing phase


RAW DATA
Poisoning
FEATURE REPRESENTATION
Evasion
CLASSIFICATION