A gépi tanulás biztonsága

2024. tavasz, választható C tárgy

Gergely Acs

CrySyS Lab, BME HIT

acs@crysys.hu

CIA

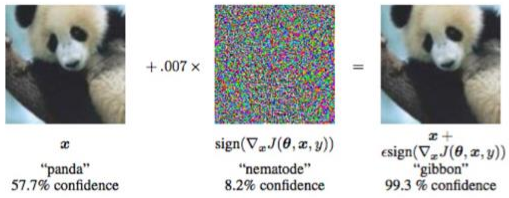
- Confidentiality

ANDY GREENBERG SECURITY 09.30.2016 11:06 AM

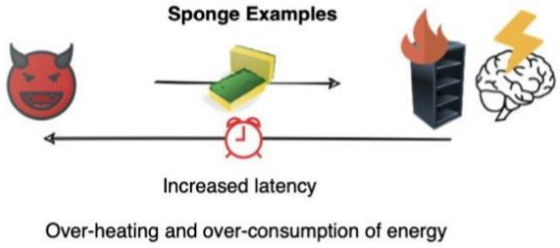
How to Steal an AI

Researchers show how they can reverse engineer and even fully reconstruct someone else's machine learning engine—using machine learning.

- Integrity



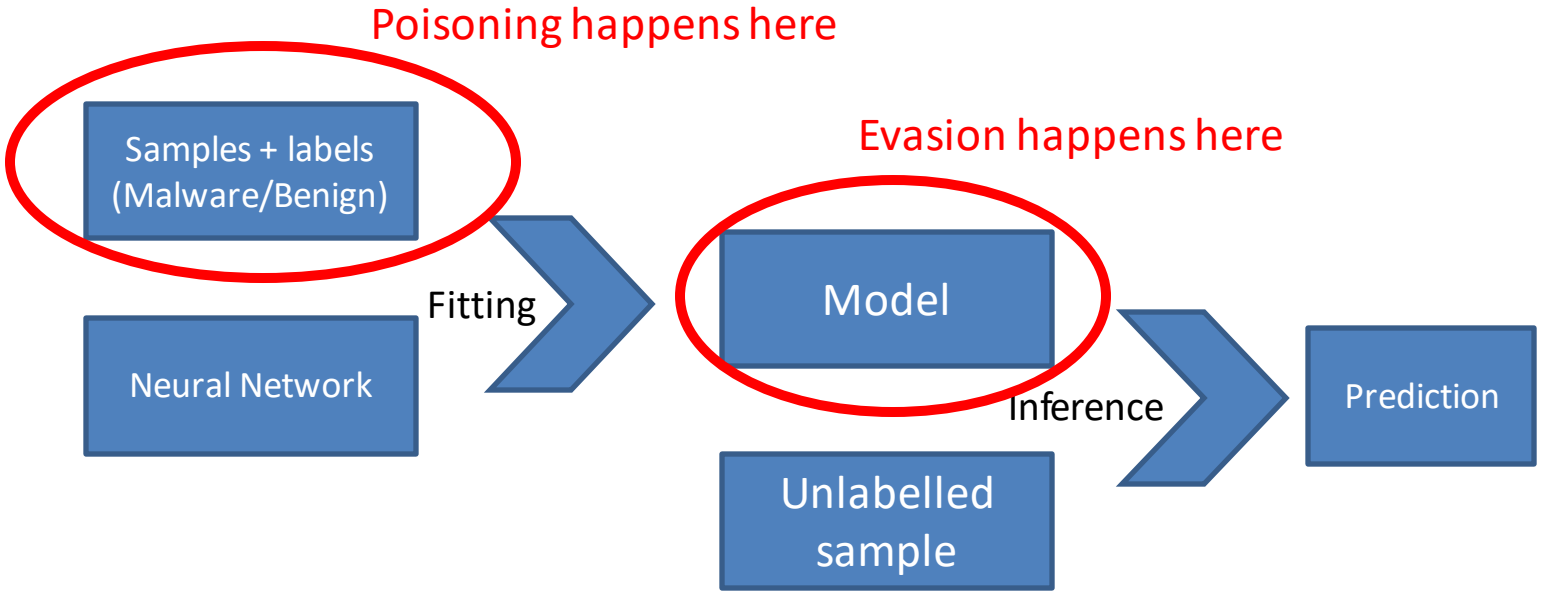
- Availability



Security of Machine Learning

- Privacy and Security are important to build **trust**
  - Safety-critical and real-time systems rely on trust

- AI/ML systems are vulnerable to many attacks at different points of the machine learning pipeline



- Security and privacy **audit of AI** are mandated by different regulations, companies cannot overlook and need experts

Requirements

- 2 lectures per week
- 1 mid-term test
- 1 homework
- 1 (written) exam
- 6 laboratory exercises (bi-weekly)

- for **5 credits**

Diploma projects

- Federated learning security
- Fairness/Robustness/Accuracy/Privacy Trade-Off
- Poisoning, adversarial examples
- Differential Privacy
- (De-)Anonymization
- Appilcations:
  - Malware detection with Machine Learning
  - Privacy-preserving processing and sharing of medical data
- …

- https://www.crysys.hu/education/projects/?q=Privacy