Machine Learning for "Big Data" in Biomedicine

Yanjun (Jane) Qi, PhD
Department of Computer Science,
University of Virginia

2016.09.16

Machine Learning is Changing the World

How may I help you, human?

Apple Siri / Amazon Echo

IBM WATSON

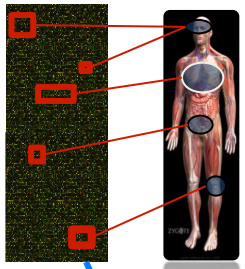
Control learning

Object recognition

Many more !

Big and Complex Data in Biomedicine

Molecular signatures of tumor / blood sample

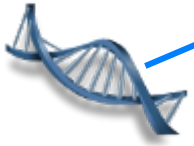


Signs & Symptoms

Patient Medical History & Demographics

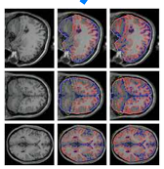
Public Health Data

Genetic Data

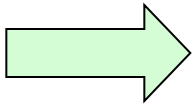


Mobile medical sensor data

Medical Images

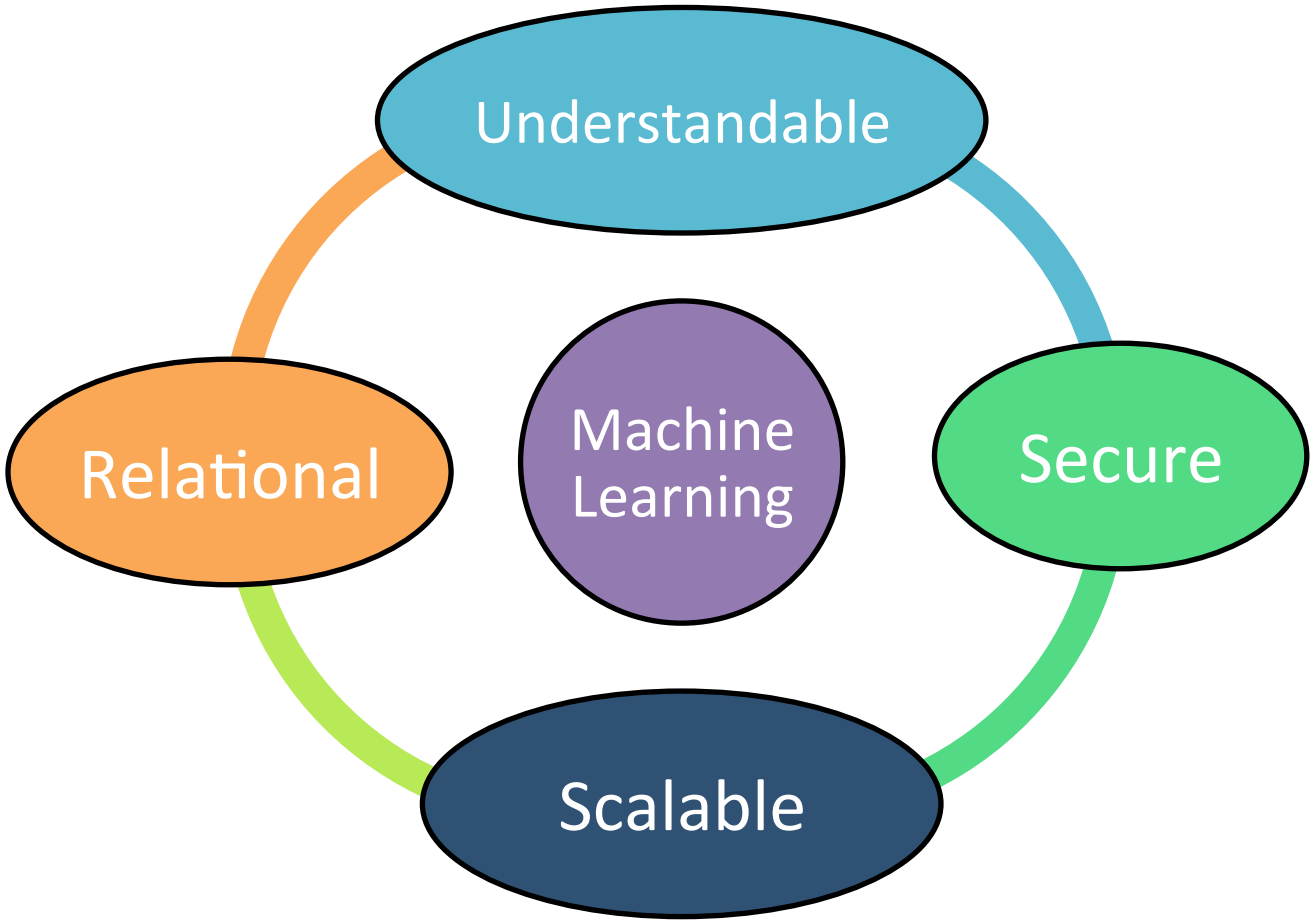


BIG Biomed DATA Analytics System



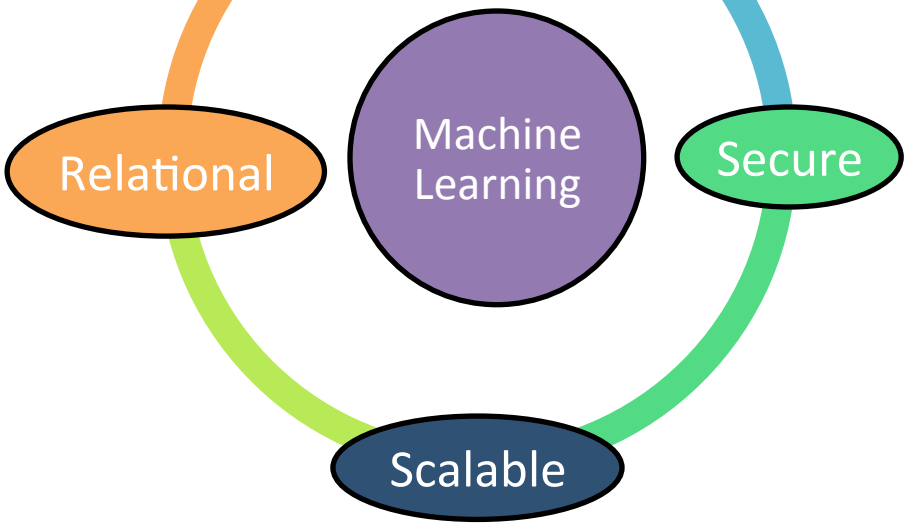
Machine learning can help in many aspects

Machine Learning for Mining "Big & Complex" Data in Biomedicine



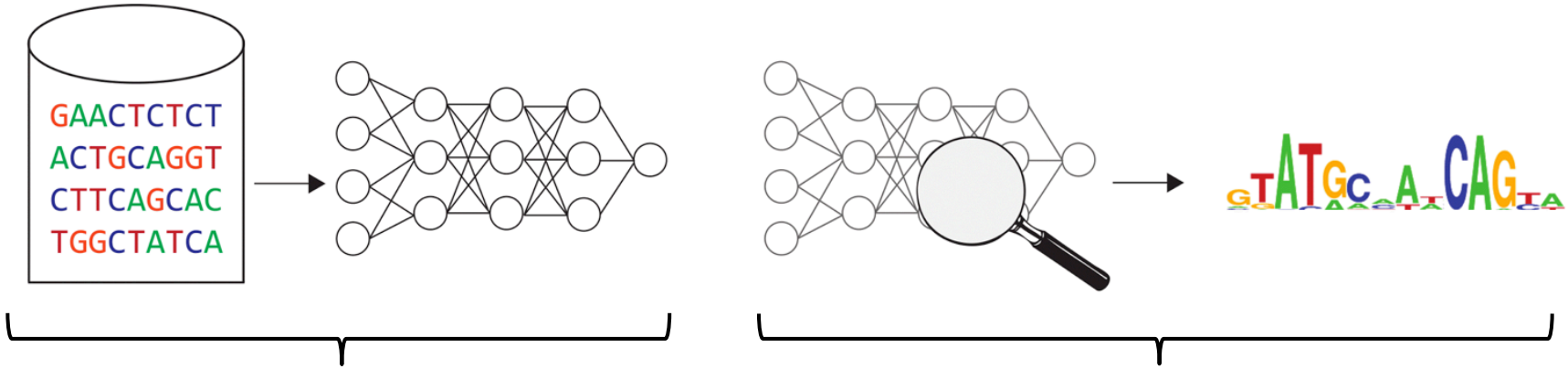
1. Understandable:

Opening Up the "Black Box" of machine learning models



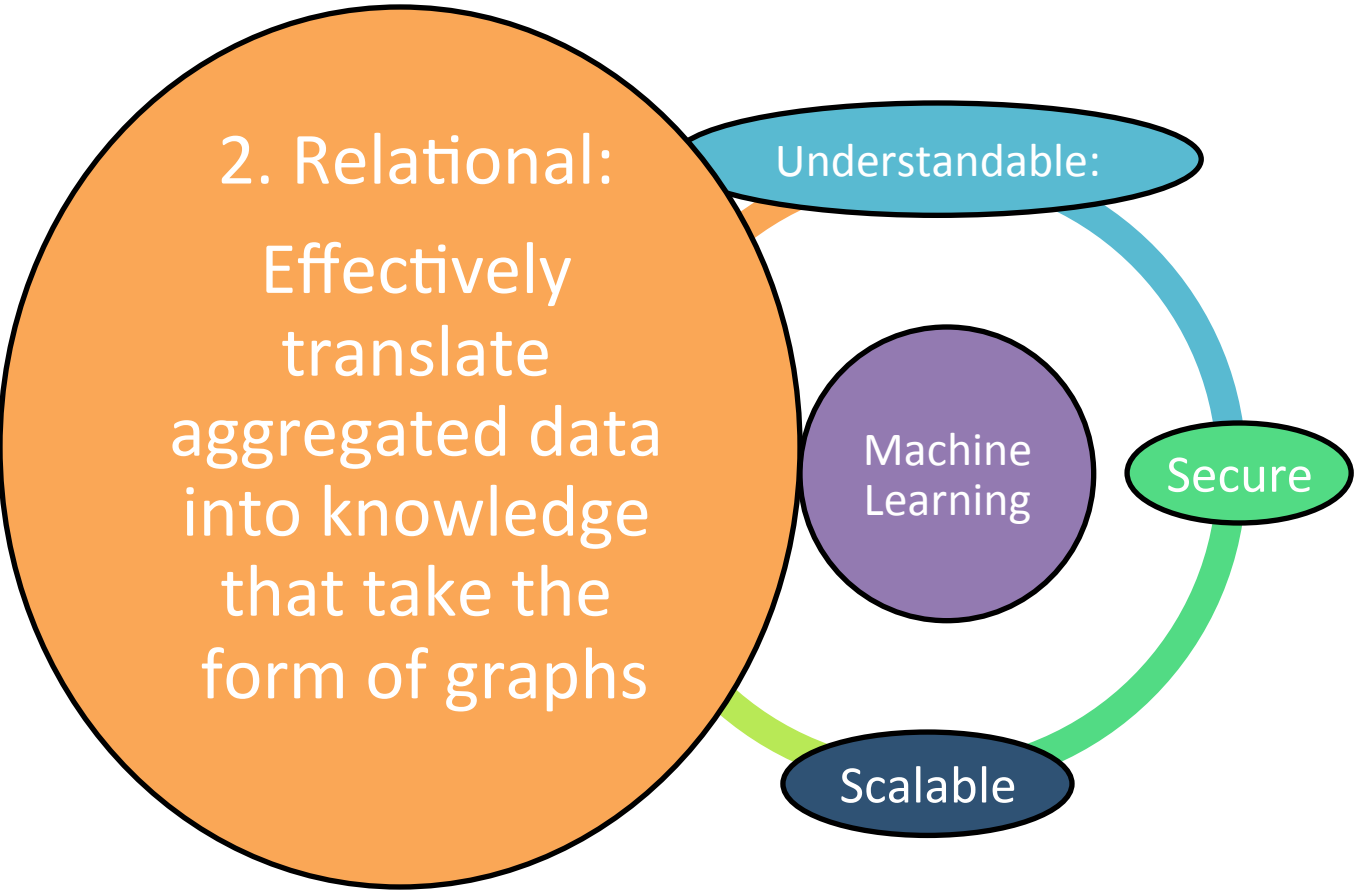
1. Understandable: Opening Up the "Black Box" of machine learning models

- Deep Genomic Dashboard: Visualizing and Understanding Genomic Sequences Through Deep Neural Networks



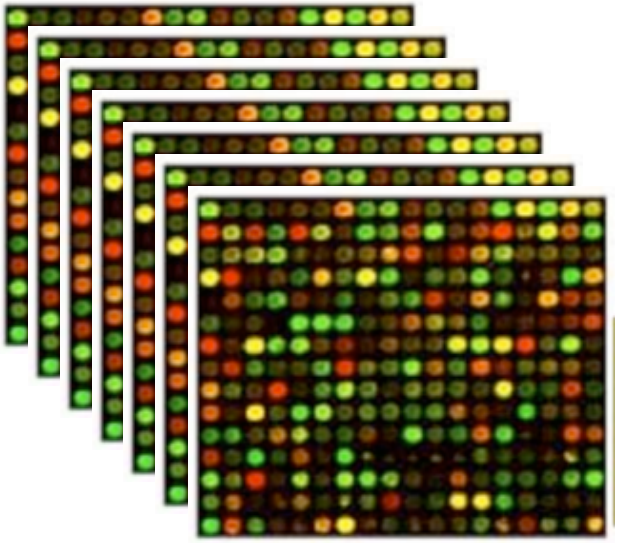
Transcription Factor Binding Site (TFBS) Prediction

Pattern (motif) Extraction


2. Relational:
Effectively translate aggregated data into knowledge that take the form of graphs
Understandable:
Machine Learning
Secure
Scalable

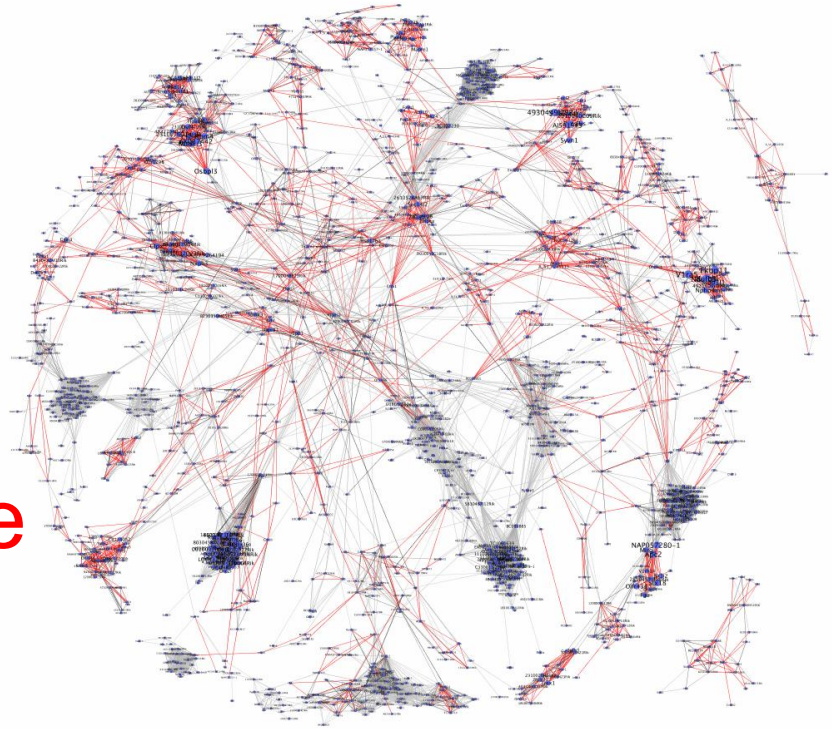
2. Relational:

Effectively translate aggregated data into knowledge that take the form of graphs



data

Inference



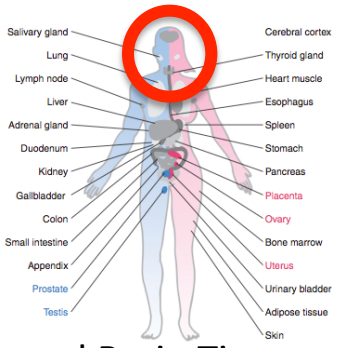
graph

Joint Inference of **Multiple** Related Graphs from **Heterogeneous** Data

Multi-Context

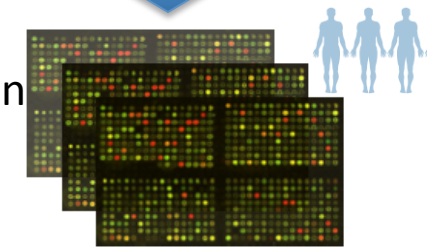
Brain Cancer

Lung Cancer

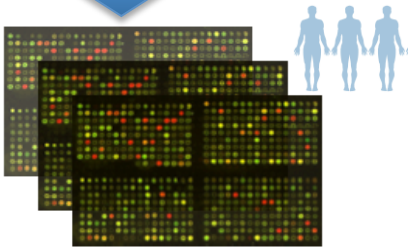


Normal Brain Tissue

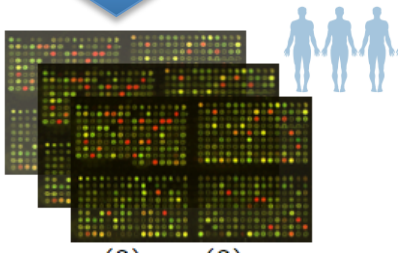
Expression Data



$(X_1^{(1)}, \ldots, X_{n_1}^{(1)}), X_i^{(1)} \in \mathbb{R}^p$



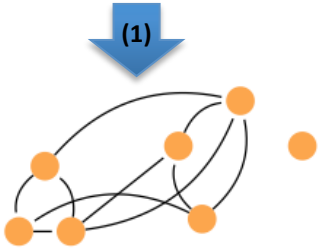
$(X_1^{(2)}, \ldots, X_{n_2}^{(2)}), X_i^{(2)} \in \mathbb{R}^p$

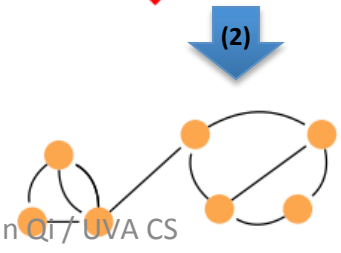


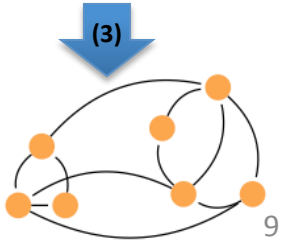
$(X_1^{(3)}, \ldots, X_{n_3}^{(3)}), X_i^{(3)} \in \mathbb{R}^p$

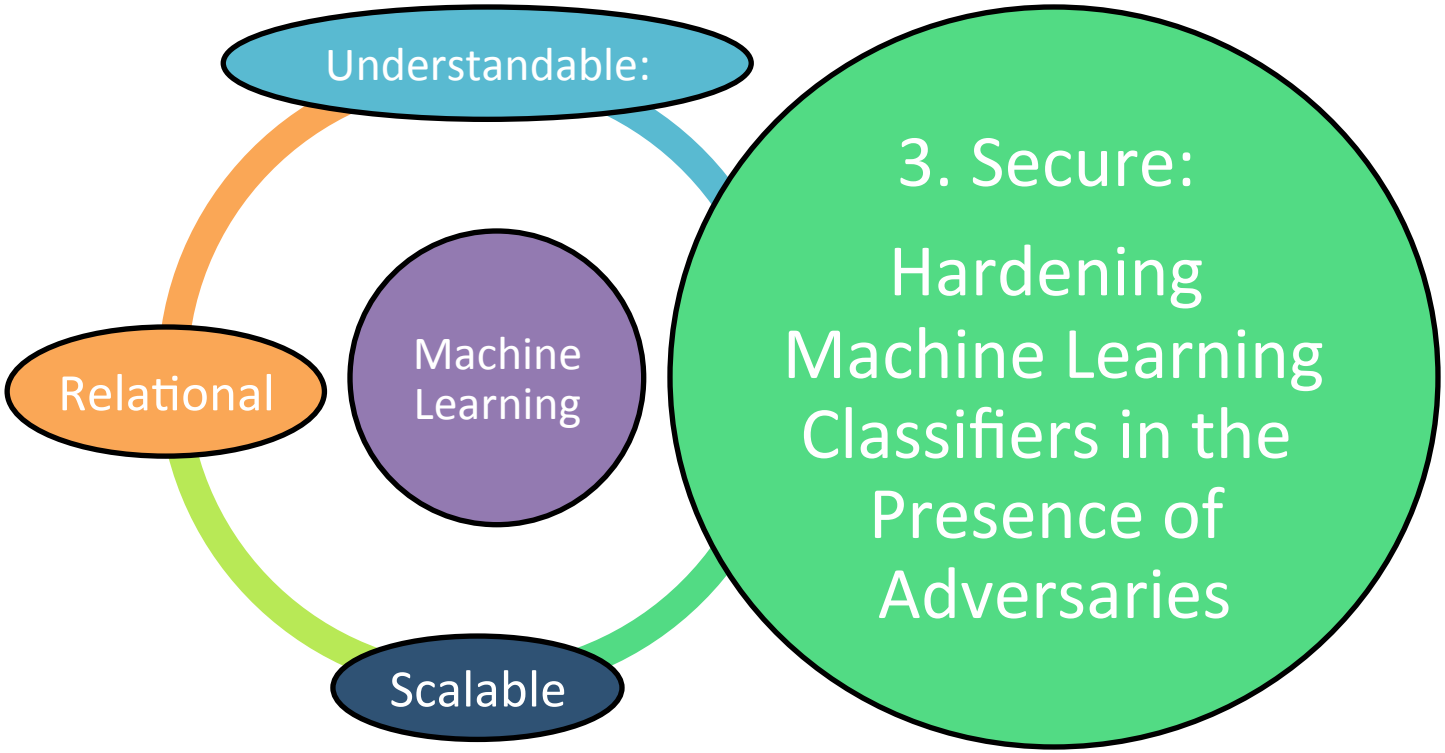
inference

Network Inference








Understandable:
Relational
Machine Learning
Scalable
3. Secure:
Hardening Machine Learning Classifiers in the Presence of Adversaries

3. Secure:

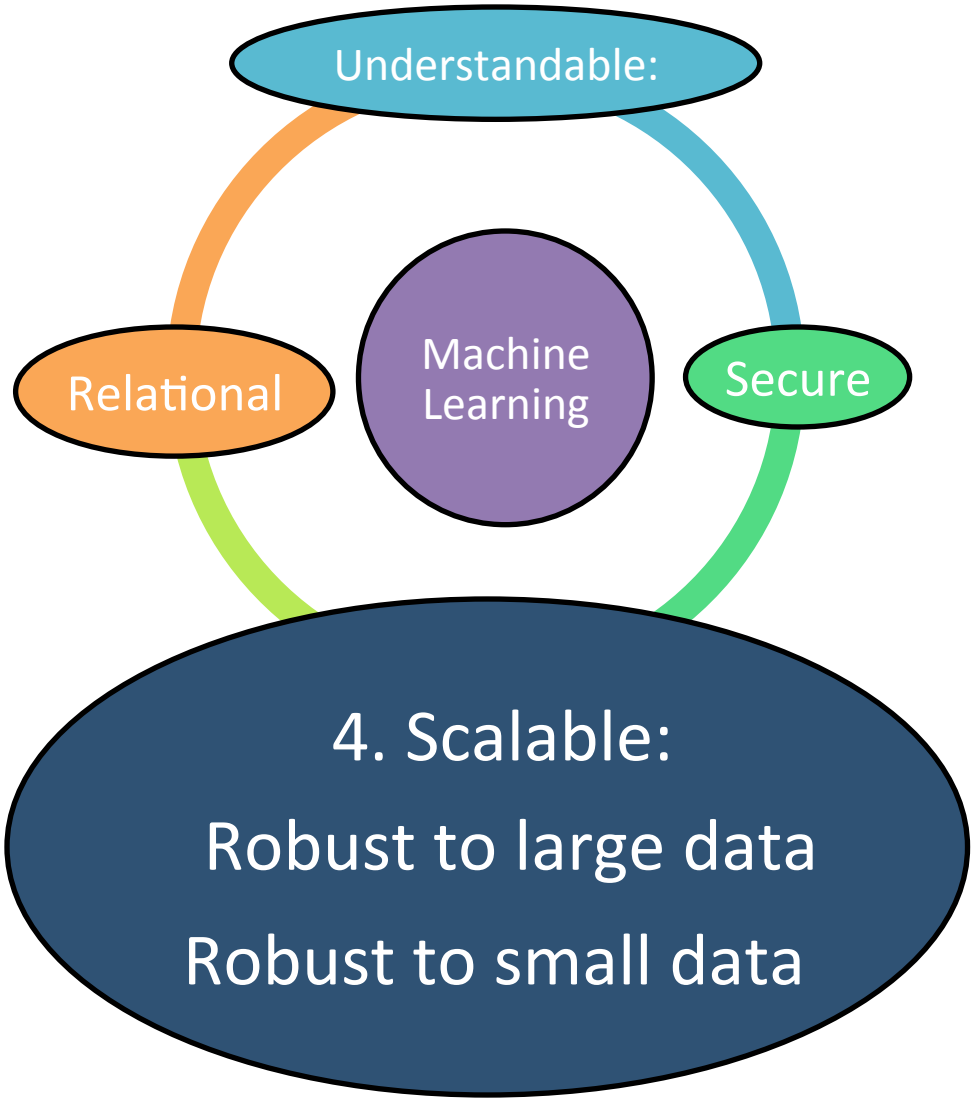
Hardening Machine Learning Classifiers in the Presence of Adversaries

- Automatically Evading Classifiers

The presence of **adversary**, with evolving evasion attacks.

Goal: Understand how ML models work under attack.

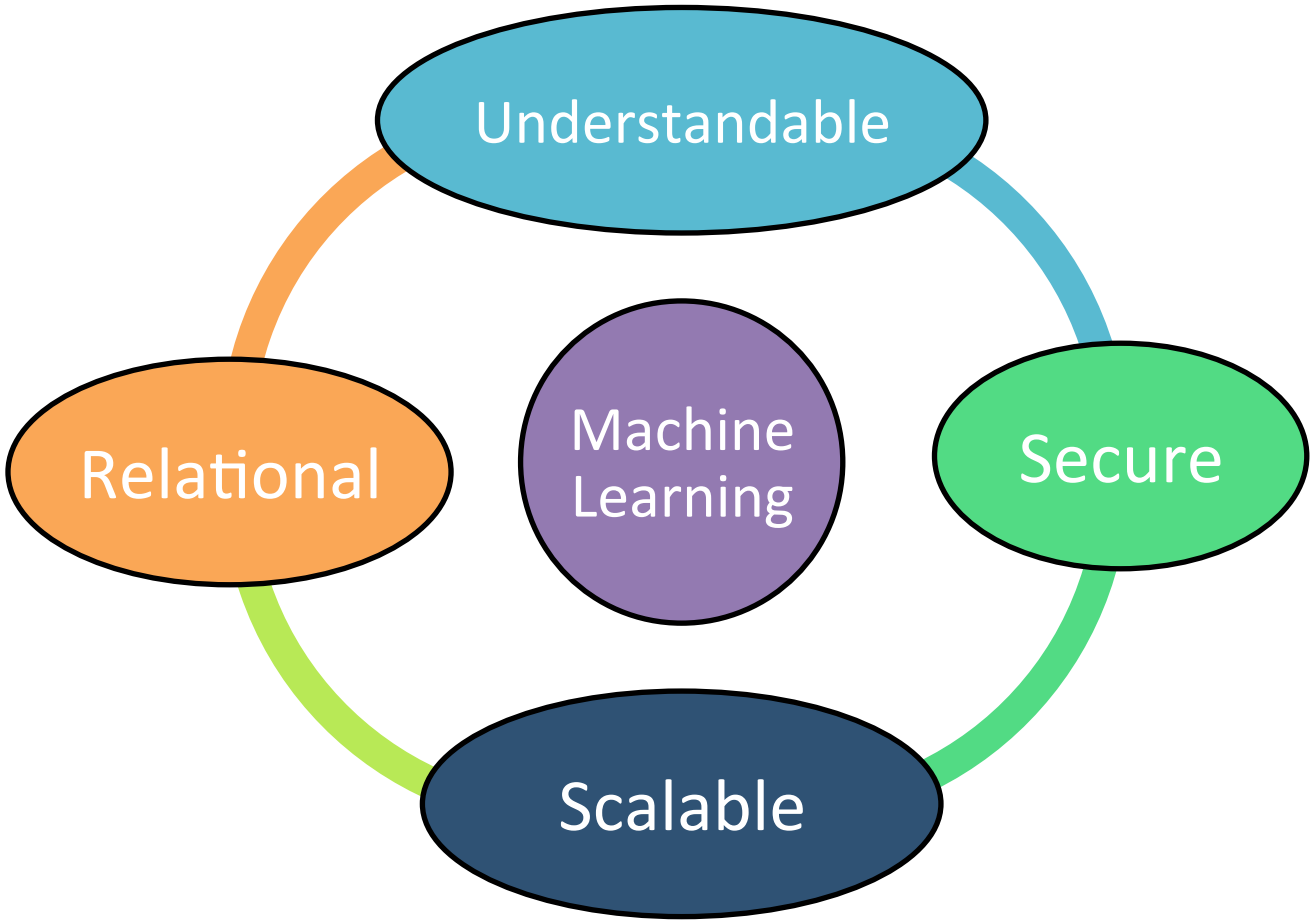
Results: Powerful machine learning models are **vulnerable**.


Understandable:
Relational
Machine Learning
Secure
4. Scalable:
Robust to large data
Robust to small data

4. Scalable:

Robust to large data

Robust to small data

- MUST-CNN: A MUltilayer Shift-and-sTitch Deep Convolutional Architecture for Sequence-based Input Output Prediction
  - Speed is key - predictions for half million samples in under two seconds

Machine Learning to Change "Big & Complex" Data in Biomedicine