Big Data for Reliability and Security

Course Syllabus

Saurabh Bagchi

PURDUE UNIVERSITY

Last update: *October 2020*

This is a course suitable for Masters and PhD students. It briefly covers the theoretical aspects of big data for reliability and security and stresses the practical systems aspects of such techniques. There are two challenge programming problems based on large real-world datasets that we have collected and curated.

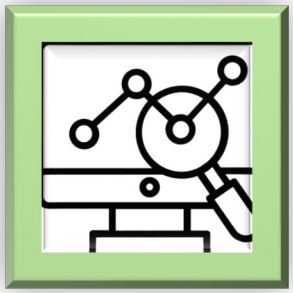
5 weeks = 15 hours of lecture and lab material.

Prerequisites: Python programming, basic knowledge of probability and statistics.

A. Foundational material on reliability and security [2 lectures]

1. Introduction: Motivation, System view of reliable and secure design, Terminology *[0.5 lecture]*
2. Reliability landscape for connected systems: Traditional concerns, new concerns due to large-scale systems, new concerns due to big data *[0.5 lecture]*
3. Security landscape for connected systems: Traditional threats, new threats due to large-scale systems, new threats due to big data *[1 lecture]*

B. Data analytic techniques for dependability [5 lectures]



1. Supervised and unsupervised learning techniques *[1 lecture]*
2. Neural Networks building blocks *[1 lecture]*

3. Techniques for dealing with large scale data: regularization, feature engineering, dimensionality reduction, etc. *[1 lecture]*
4. What is our toolchest of data analytic techniques: what to use and when *[1 lecture]*
5. Data analytic techniques used for reliability and security: strengths, weaknesses, opportunities *[1 lecture]*

C. Big data security and insecurity [5 lectures]

1. Attacks against big data algorithms: evasion and poisoning attacks *[1 lecture]*
2. White box and black box attacks *[1 lecture]*
3. Defenses: Adversarial training, defensive distillation, adversarial example detection *[1 lecture]*
4. Machine learning at scale: Federated learning *[1 lecture]*
5. Federated learning for security and privacy *[1 lecture]*

D. Case studies and challenge problems [3 lectures]

1. Case studies on adversarial Machine Learning: Image and video manipulation *[1 lecture]*
2. Systems for big data processing: Spark, TensorFlow on cluster, TensorFlow Light. Benchmarks for big data processing *[1 session: hybrid of lecture and lab]*
3. Challenge problems *[1 session: hybrid of lecture and lab]*
    a. Challenge problem 1: Predicting computer system failures
    b. Challenge problem 2: Proximity detection through Bluetooth signals