Using Machine Learning for Network Capacity Management

Speaker: Taghrid Samak
Host: Lori Pollock

CRA-W Undergraduate Town Hall
November 9th, 2017

CRA-W
Computing Research Association
Women

Speaker & Moderator

Taghrid Samak

Taghrid Samak holds a doctorate degree in computer science from DePaul University, a BSc and MSc in computer science from Alexandria University in Egypt, and is currently pursuing her Juris Doctorate degree at the University of San Francisco. At Google, Taghrid applies statistical modeling for diverse network applications from capacity planning to wireless networks. Previously, she worked at Lawrence Berkeley National Laboratory where her research focused on applying data analysis and machine learning to enable cross-discipline scientific discovery.

Taghrid is co-founder and steering committee member of the Arab Women in Computing organization and volunteers as a mentor for various women in computing organizations.

Lori Pollock

Dr. Lori Pollock is a Professor in Computer and Information Sciences at University of Delaware. Her current research focuses on program analysis for building better software maintenance tools, software testing, energy-efficient software and computer science education. Dr. Pollock is an ACM Distinguished Scientist and was awarded the University of Delaware's Excellence in Teaching Award and the E.A. Trabant Award for Women's Equity.

CRA-W
Computing Research Association
Women

About Me

Background

- Originally from Alexandria, Egypt
- BSc and MSc Computer Science, Alexandria University, Egypt
- PhD in Computer Science, DePaul University
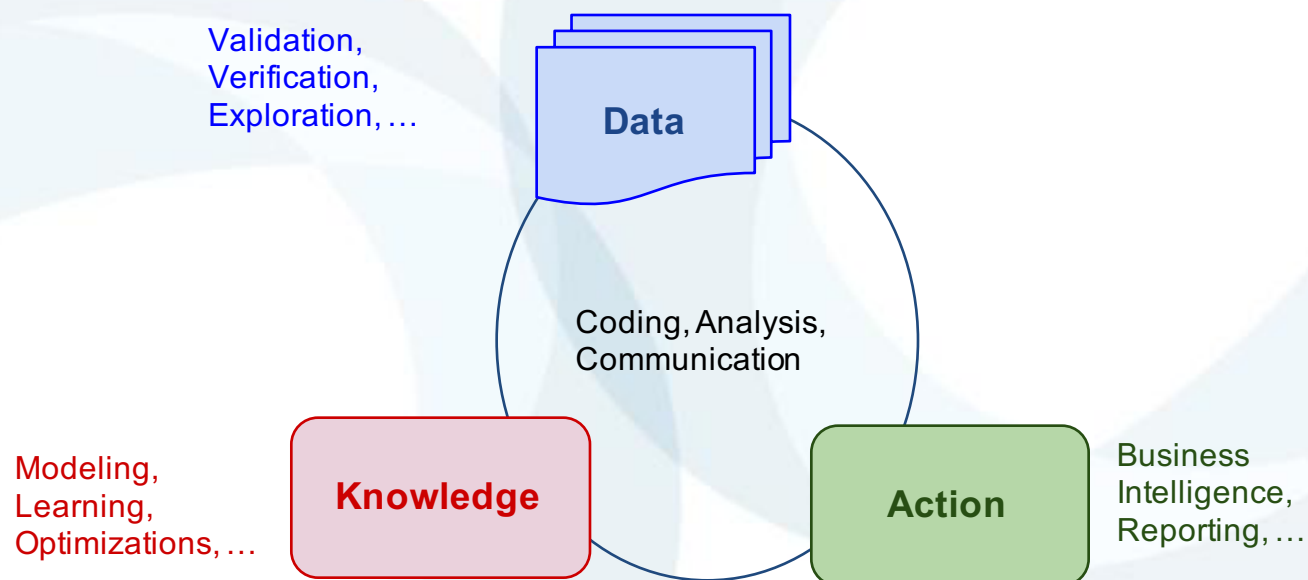- JD Student, University of San Francisco School of Law

Career

- *Currently:* Sr. Data Analyst @ Google, Corporate Networking

- Research Scientists @ Lawrence Berkeley National Lab (data analysis for Biology, Physics, Systems, …)
- Research Intern @ Bell Labs
- Research Assistant @ DePaul University
- Teaching Assistant @ Alexandria University

CRA-W
Computing Research Association
Women

"A Day in the Life....of a Data Scientist"

Agenda

- Background
  - Capacity planning in enterprise networks
  - Machine learning

- Usage forecast for Google's enterprise network
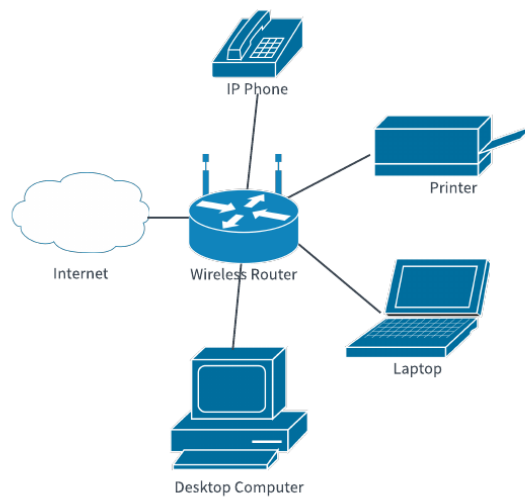 - Data
  - Model - knowledge
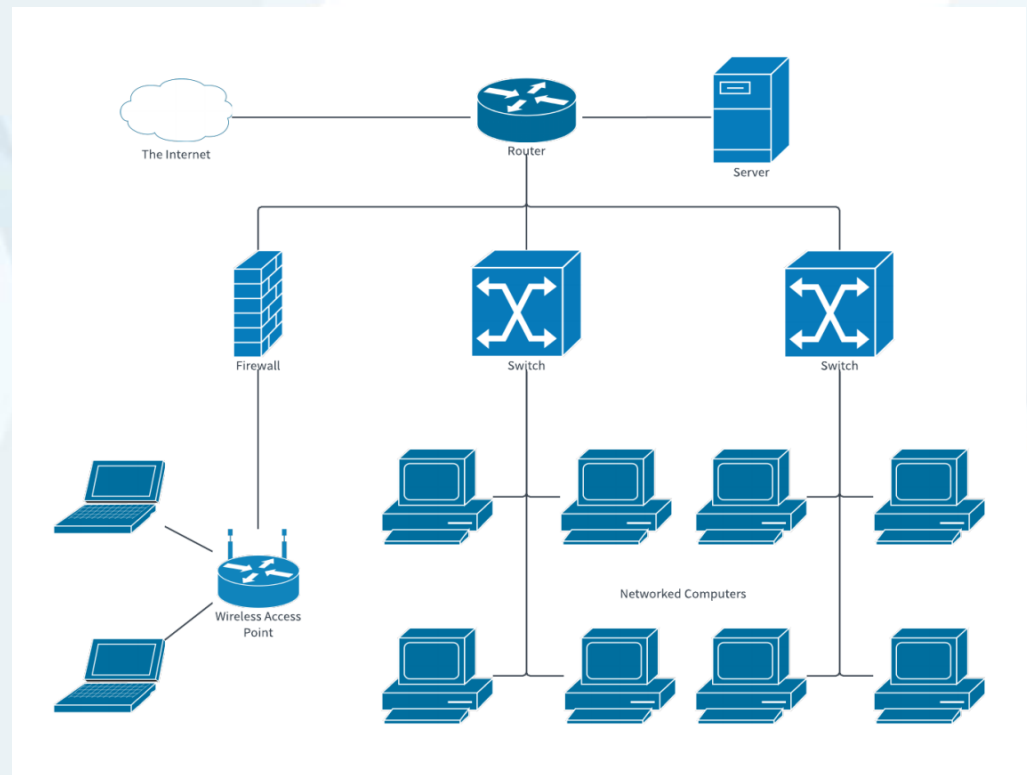 - Results - action

CRA-W
Computing Research Association
Women

Networking Overview

Home Network

IP Phone
Printer
Internet
Wireless Router
Laptop
Desktop Computer

Office Network

The Internet
Router
Server
Firewall
Switch
Switch
Wireless Access Point
Networked Computers

Images from:
https://www.lucidchart.com/pages/examples/network-diagram

CRA-W
Computing Research Association
Women

Network Capacity Management

- Ensuring sufficient resources on the network to satisfy performance requirements

- Home network
  - choosing subscription from the Internet Service Provider

- Enterprise Network
  - interconnected office networks
  - office design optimizations
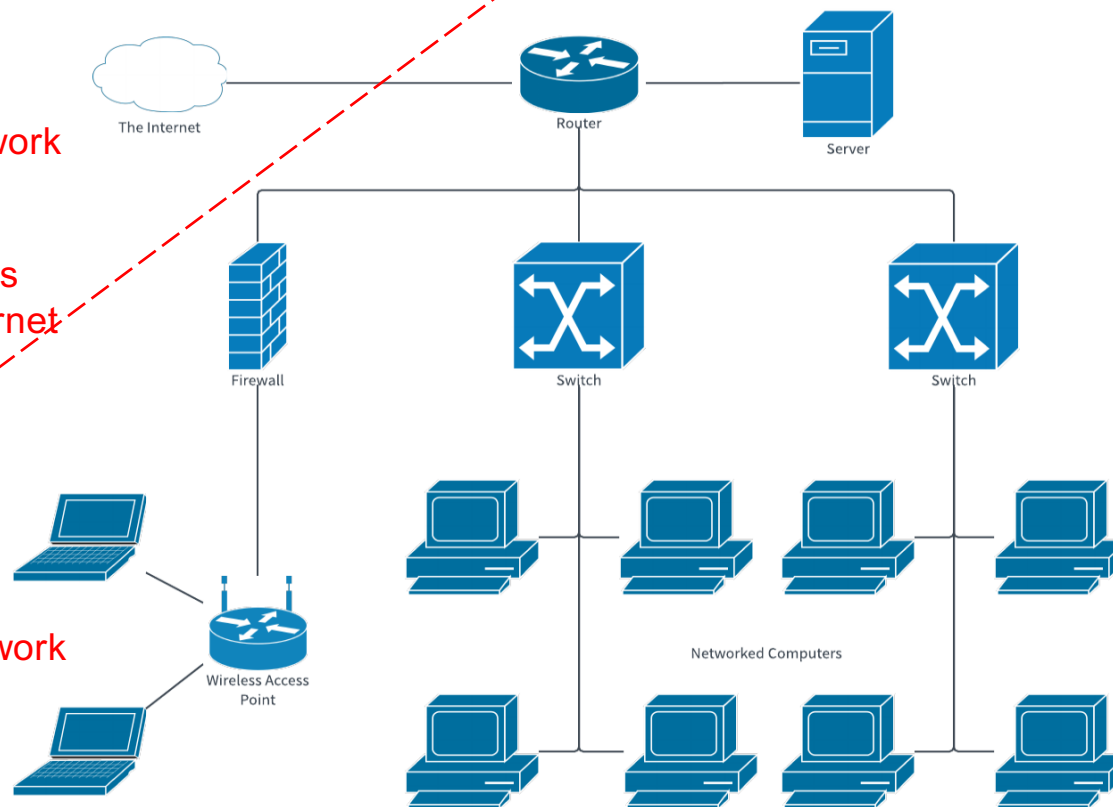 - managing bandwidth inside and outside of the enterprise

Capacity Management Points

Wide Area Network
WAN

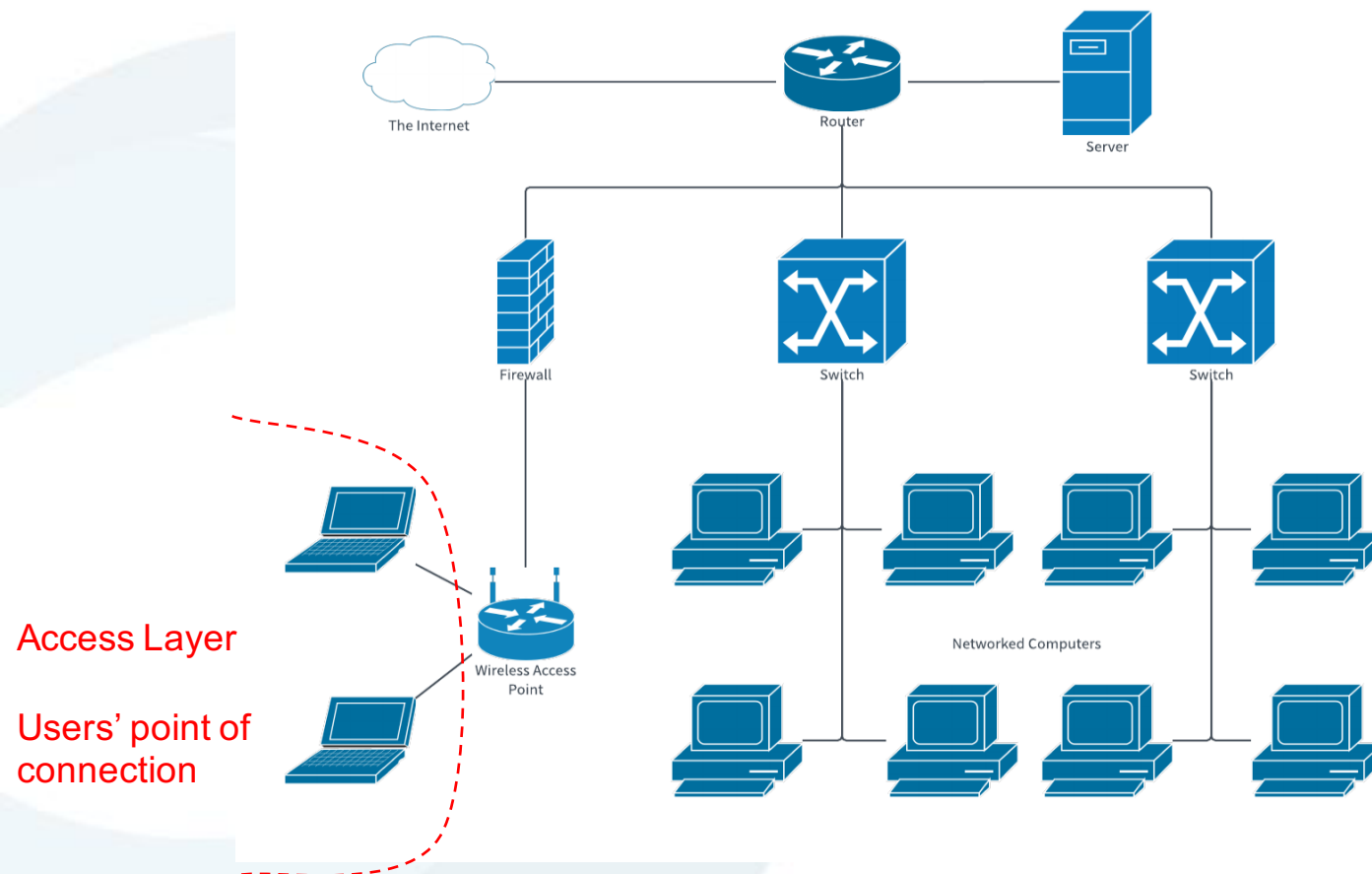- Multiple Offices
- Offices to Internet

Local Area Network
LAN

- Within Offices

The Internet

Router

Server

Firewall

Switch

Switch

Wireless Access
Point

Networked Computers

CRA-W
Computing Research Association
Women

Capacity Management Points

The Internet

Router

Server

Firewall

Switch

Switch

Access Layer

Users' point of connection

Wireless Access Point

Networked Computers

CRA-W
Computing Research Association
Women

Enterprise Network Data Analysis

- Data
  - Traffic passing through the network at each level from access layer to WAN

- Knowledge → Action
  - Which users or applications are using the network?
  - Can we optimize the network design?
  - When do we need capacity upgrade or downgrade for a specific office?
  - Can we predict performance problems?

What is Machine Learning?

"Field of study that gives computers the ability to learn without being explicitly programmed"
- wikipedia definition

"How can we build computer systems that automatically improve with experience, and what are the fundamental laws that govern all learning processes?"
- Tom M. Mitchell, CMU

The Data

- Table of observations/samples/objects/…
- Each observations has dimensions/attributes/ features/fields/variables/measurements/…

	Feature 1	Feature 2	Feature 3	…	Feature k
Observation 1					
Observation 2					
…					
Observation n					

Typical Workflow

- Data Collection
- Preprocessing (filtering, scaling, sampling, …)
- Feature Extraction
- Dimensionality Reduction
- Learning the model (build knowledge)
  - Training
  - Testing
  - Validation
- Use the model
  - predictions
  - actions

CRA-W
Computing Research Association
Women

The Learning Process

- Can we find a function that accurately fits the data?

- Supervised learning
  - the function predicts a feature of interest accurately for the training data and new samples
- Unsupervised learning
  - the function creates the correct pattern/groups of data

- What's needed
  - Data
  - Hypothesis space of potential functions
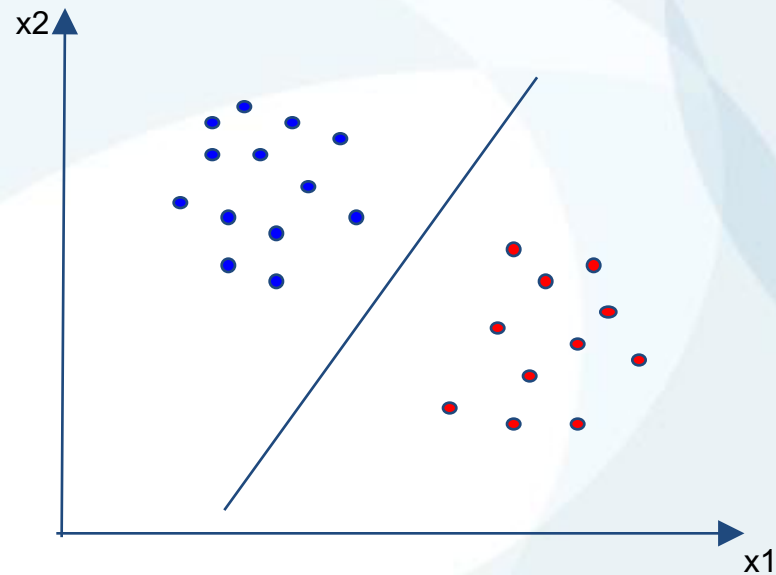 - Optimization function to minimize the error
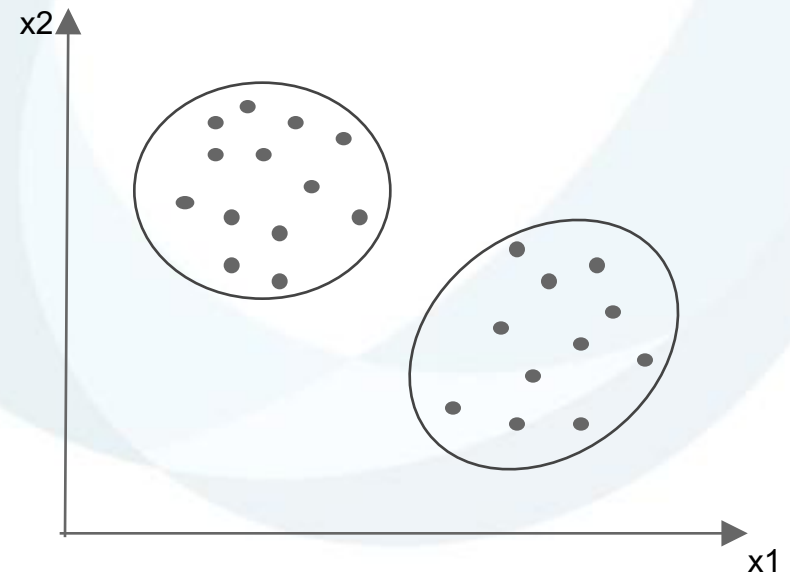
CRA-W
Computing Research Association
Women

Machine Learning Methods

Supervised

Unsupervised

Supervised Learning

- Independent variables (X): dimensions/features/…
- Dependent variables (Y): classifications/labels/…
- Learning Y's from values of X's
- f: X → Y
- Usually only one "y"

	x1	x2	x3	y1	y2
Observation 1					
Observation 2					
…					
Observation n					

Labels

CRA-W
Computing Research Association
Women

Unsupervised Learning
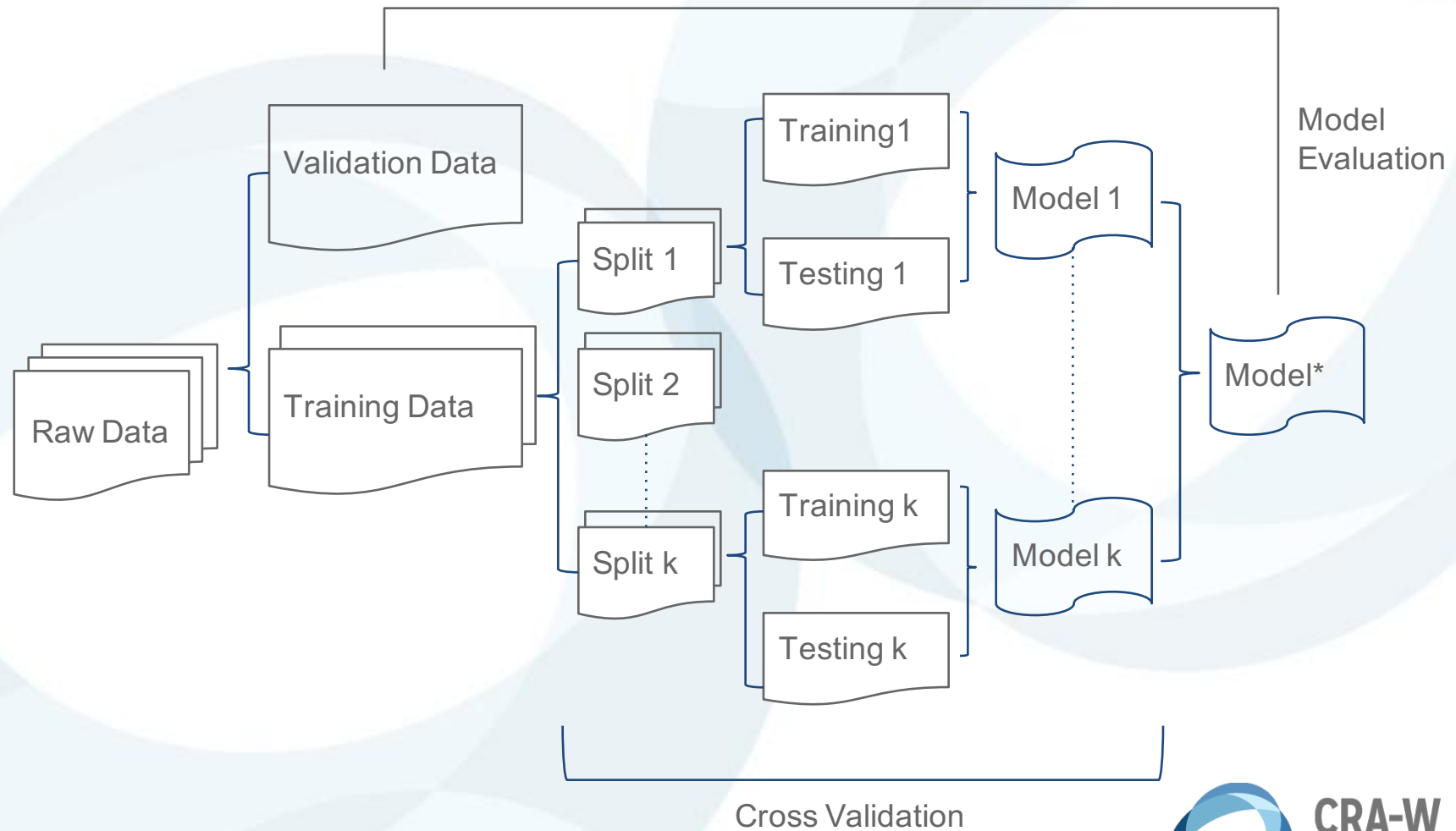
- No labels
- Learning "patterns" from values of X's

	x1	x2	x3	...	xk
Observation 1					
Observation 2					
...					
Observation n					

CRA-W
Computing Research Association
Women

The Learning Process - Data Flow

Raw Data → Validation Data, Training Data

Training Data → Split 1, Split 2, ... Split k

Split 1 → Training1, Testing 1 → Model 1

Split k → Training k, Testing k → Model k

Model 1 ... Model k → Model*
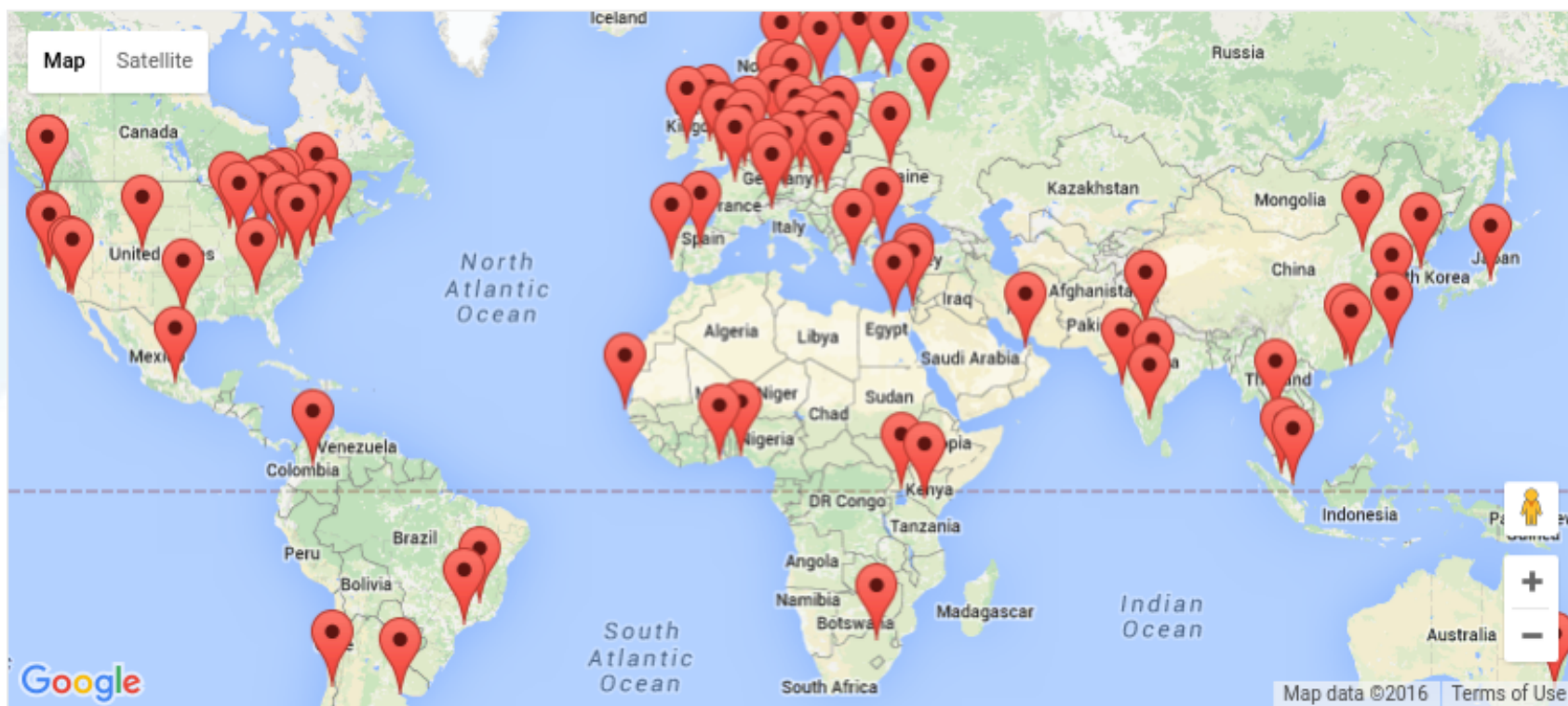
Model Evaluation

Cross Validation

CRA-W
Computing Research Association
Women

Google Offices WAN Capacity Forecast

- Forecasting network usage for each office based on historical data
  - When do we need circuit capacity upgrade or downgrade for a specific office?
  - How to model usage changes for changing headcounts?

Taghrid Samak, Mark Miklic, "WAN Capacity Forecasting for Large Enterprises."
IEEE/IFIP International Workshop on Analytics for Network and Service Management,
AnNet 2016

CRA-W
Computing Research Association
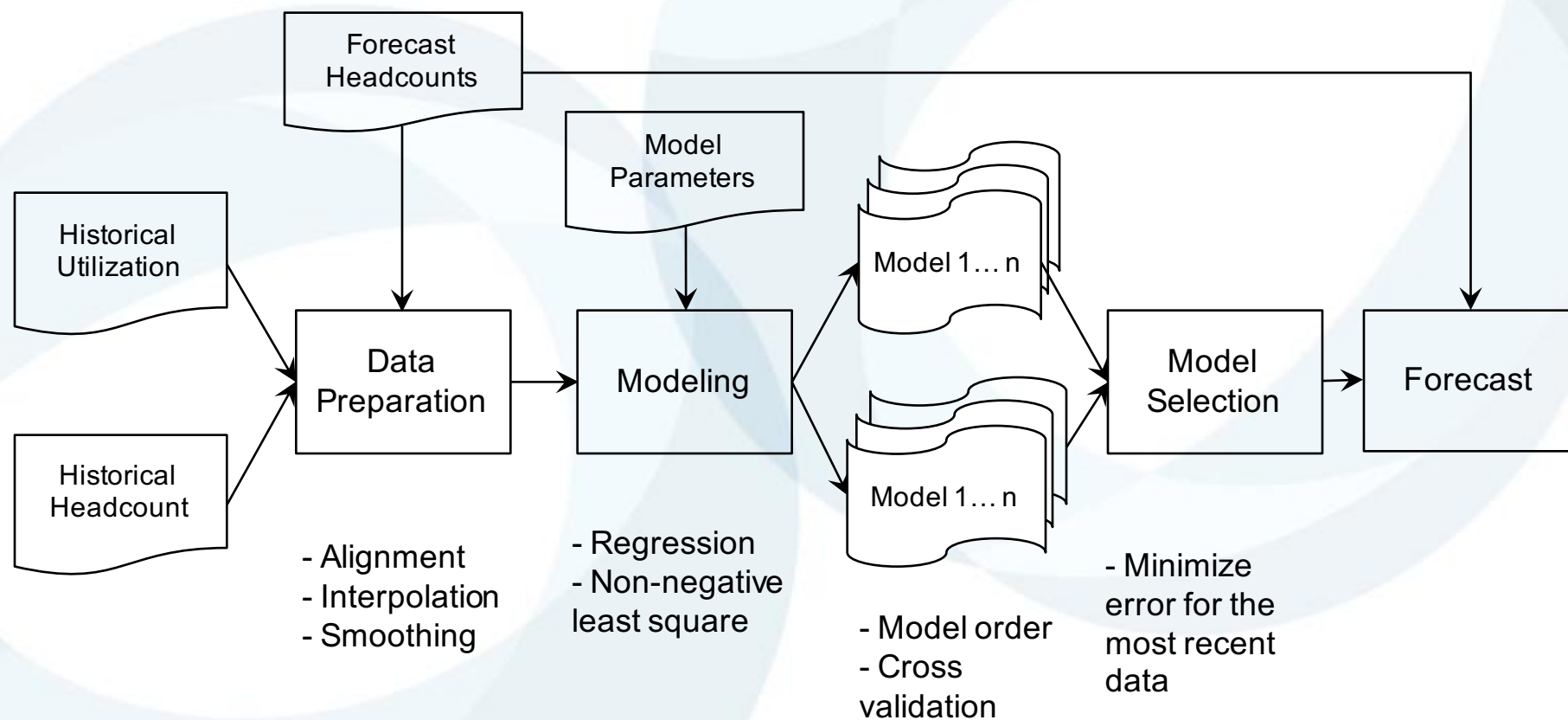Women

Google Enterprise Network

Data

- Historical inbound/outbound bandwidth utilization for each office - SNMP
- Historical and forecast headcount - HR
- $BW = f(hc, T)$
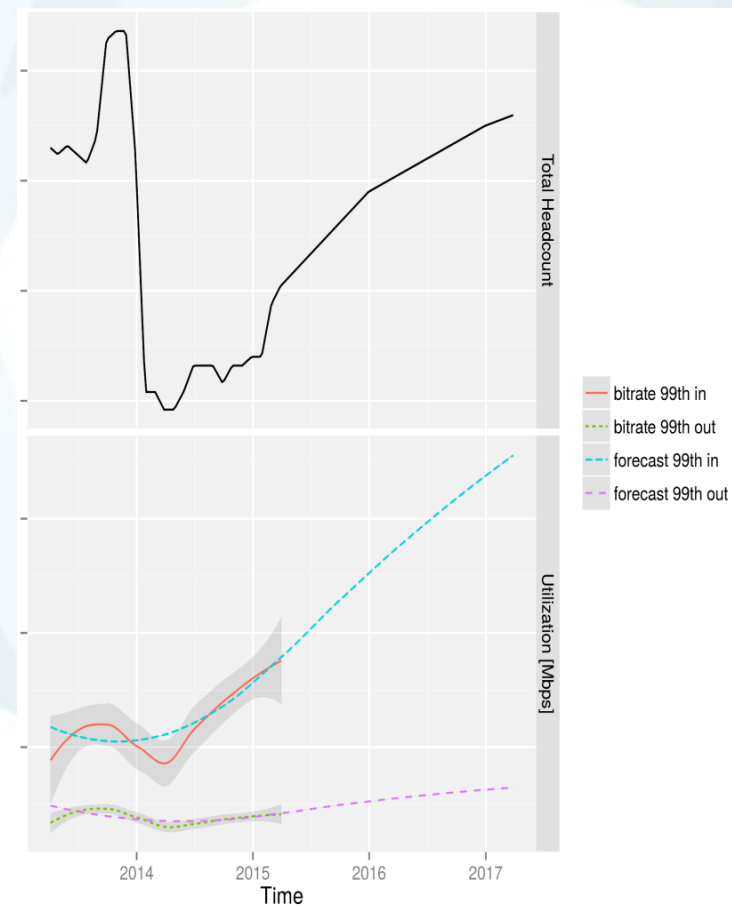
- Approximately 400K samples per office

Modeling Process Per Office

Forecast Results
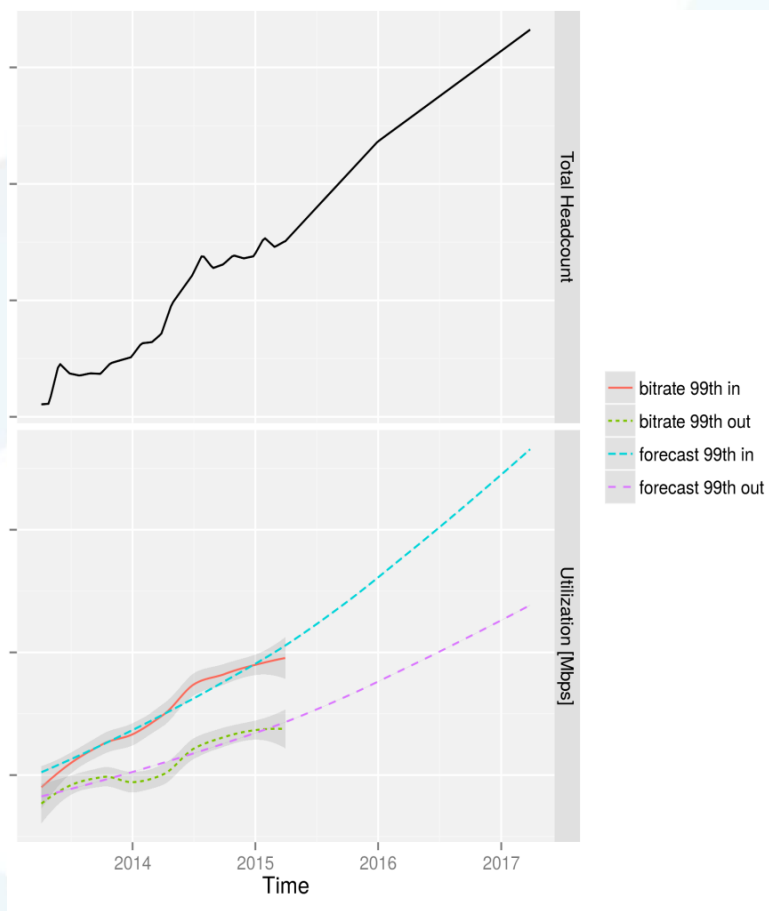
Extracurricular Activities and Time Management

Speaker: Taghrid Samak
Host: Lori Pollock

CRA-W
Computing Research Association
Women

Before we start

- What works for you might not work for everyone
- Find your own pace and balance

- Here is some anecdotal advice :)

CRA-W
Computing Research Association
Women

Taghrid's Extracurriculars

- As a graduate student
    - UPE Honor Society, DePaul Chapter President
    - ACM-w, DePaul Chapter Treasurer
    - Egyptian Student Association National Executive Committee
- As a professional
    - Program committee member for technical conferences
    - Arab Women in Computing Co-founder (arabwic.org)
    - Mentor for US Dept of State Techwomen Program (techwomen.org)
    - Internal diversity efforts within Google
    - USF Law School Dean's Merit Scholarship

CRA-W
Computing Research Association
Women

Extracurricular Activities

- Activities outside of the "normal" realm of study/work

- As a student
  - Normal → school work, part time job
  - Extracurricular → student organizations, sports, non-profit, …

- As a professional
  - Normal → full time job
  - Extracurricular → sports, non-profit, mentoring, …

CRA-W
Computing Research Association
Women

Extracurricular Activities

- Which activity is right for me?
  - Follow your passion
  - Commit and follow through

- Personal benefits
  - Helping causes you care about
  - Building friendships

- Professional benefits
  - Building resume and network
  - Learning new skills
  - Getting experience

CRA-W
Computing Research Association
Women

Time Management Strategies - prep work

- Clear and focused goals/tasks
  - Your to-do list
  - Dynamic and flexible
  - Learn to say "no"

- Prioritize
  - By value
  - By time needed
  - By deadline

CRA-W
Computing Research Association
Women

Time Management Strategies - steps

- Planning
  - By priority
  - Day-to-day
  - Short- and long-term
- Execution
  - Avoid procrastination
  - Limit interruptions
  - Limit multitasking
- Evaluate and readjust
  - Identify areas of high/low productivity
  - Redefine priorities
  - Ask for help

CRA-W
Computing Research Association
Women