ECE/CS 498 DS - Section U/G (R. Iyer)

- **Meeting time and Place:**
  - 12:30 PM – 01:50 PM, Mondays and Wednesdays, 1310 DCL
  - Discussion Section / additional TA office hours: 4:00 PM - 5:00 PM Friday (location CSL 141)
  - All class-related information will be announced through Piazza
- **Instructor:**
  - Professor Ravi K. Iyer
  - Office: 255 Coordinated Science Laboratory (CSL)
  - Email: rkiyer@illinois.edu
  - Office Hours: 11:15 AM – 12:15 PM, Mondays
  - (Other times by appointment. Please contact Heidi Leerkamp: leerkamp@illinois.edu)
- **Teaching Assistants:**
  - Yuming Wu (ywu112), Krishnakant Saboo (ksaboo2), Chang Hu (changhu), James Cyriac (jcyriac2)
  - Office Hours: Monday and Wednesday 4-5 pm at CSL 239

ECE/CS 498 DS - Section U/G (R. Iyer)

- **Recommended Text:**
  - **Class Notes and Lecture Slides**
  - Trevor Hastie, Robert Tibshirani, and Jerome Friedman, "The Elements of Statistical Learning: Data Mining, Inference, and Prediction"
- **Further Reading and Sample Problems:**
  - Daphne Koller and Nir Freidman, "Probabilistic Graphical Models: Principles and Techniques"
  - Ravi Iyer's ECE 313 Class Notes (link)
- **Class Website:**
  - **https://courses.engr.illinois.edu/ece498dsu/sp2019/**
  - **All class-related information will be announced through Piazza**
  - **A detailed class schedule including topics covered and reading for each class is on the class website; we will adjust the schedule as needed**
  - **The classes are being recorded and will be available through class website**
  - **Lecture notes will be posted on the class website weekly**
  - **Piazza - https://piazza.com/illinois/spring2019/ececs498ds/home**

Grading Policies

Activity	Grade
Mini-Projects 1, 2, 3	50% (15%, 15%, 20% each)
Midterm and Final	30%
Final Project (**Graduate Students Only**)	30%
Class Participation	10%
Homework	10%

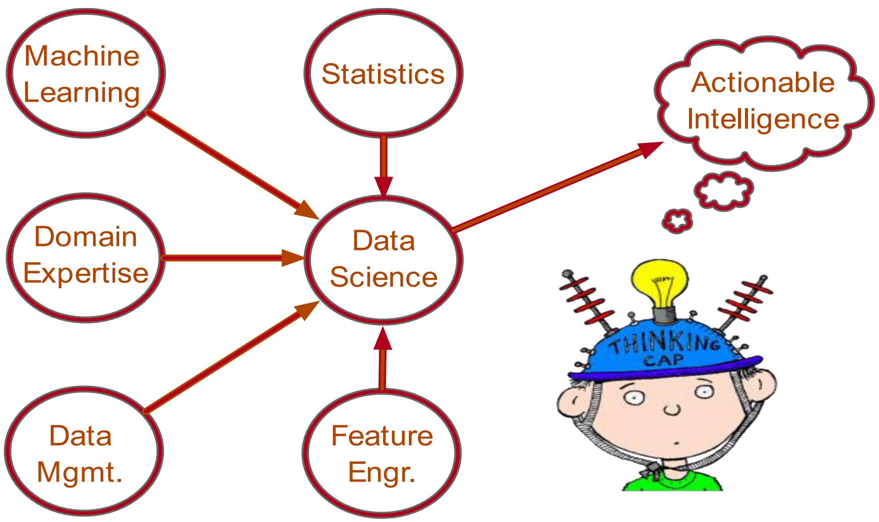
- **Undergraduate (3 hours): 100%**
- **Graduate (4 hours): 130%**
  - Will be normalized to 100%

- **There will be three mini-projects and a final project (Graduate Students Only) to provide hands-on experience with applications of data analytics/machine learning**
- **Mini-projects will be handed out or posted on the class website**
  - **Full credit for submissions on time.**
  - **Late submission policy: 10% will be taken off for every day, prorated (up to 3 days max). 0 credit after that.**
  - **Groups Policy: Students will form groups (3 persons) for the projects early in week 2, otherwise TAs will form the groups for you**
- **While we encourage discussions, submitting identical material is not allowed and will incur appropriate penalties**
- **We will hold optional discussion sessions related to the mini-projects in progress. You are strongly encouraged to attend. Location: 4pm-5pm on Fridays at CSL 141. No discussion in the first week.**
- **Project descriptions and due dates will be posted on the class web/Piazza site under Student Projects.**
  - **Please check the Resources for tutorials and hints related to projects**

Other Information

- **Prerequisites**
  - Basic probability and basic computer programming skills are essential. ECE 313 or CS 361 (Statistics and Probability), and exposure to basics of scripting languages (such as Python). Knowledge of Operating Systems (e.g., ECE 391), or an equivalent course, is beneficial.
  - HW0 to test basic probability skills, HW1 to test Python programming skills
  - **Talk to instructor if you find HW0 to be difficult**
- **Lectures (45 hours)**
  - ~30 hours classroom lectures and presentations
  - ~15 hours hands-on data analytics labs and group activities
- Use of cell phones or other non-class electronics is **NOT** allowed
- Attendance to all lectures, and in-class labs are required **(10% of the total grade)**
  - **Students can miss no more than one group activity. Note group activities are only tentatively scheduled.**
- DRES requirements must be reported to instructor/TAs by the end of 1st week

ECE 498 DS



In this course,

- **Raw datasets to actionable intelligence**
- **Building end-to-end workflows that can work in practice**
- **Work on real-world problems of high societal impact**

Solving problems in real-world involves

- Understanding the problem and the associated domain
- Finding or building relevant datasets

Use of data-science techniques to find solutions by

- Converting raw datasets to usable features
- Choosing a model that best represent the data
- Validation of the models on the field

ECE 498 DS Course Structure

Problem solving and domain-driven analyses

You will be introduced to real-world problems in domains of societal importance: (a) Safety in AVs, (b) Health Analytics, and (c) Intrusion Detection

In-class lectures, and activities

You will work on hands-on modeling tutorials with support from instructors

Homework assignments

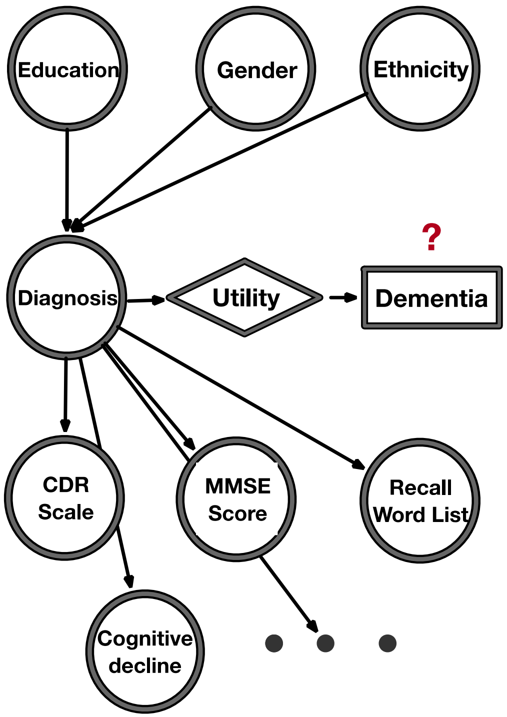
You will be given theoretical questions, and small programming assignments to strengthen your understanding the methods

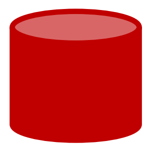
Outcomes

- ***Learn to handle the data***
- ***Learn to decide what model to use and when***
- ***Learn to apply the models to real-world examples***
- ***Learn to derive insights from by combining model solutions with domain knowledge***

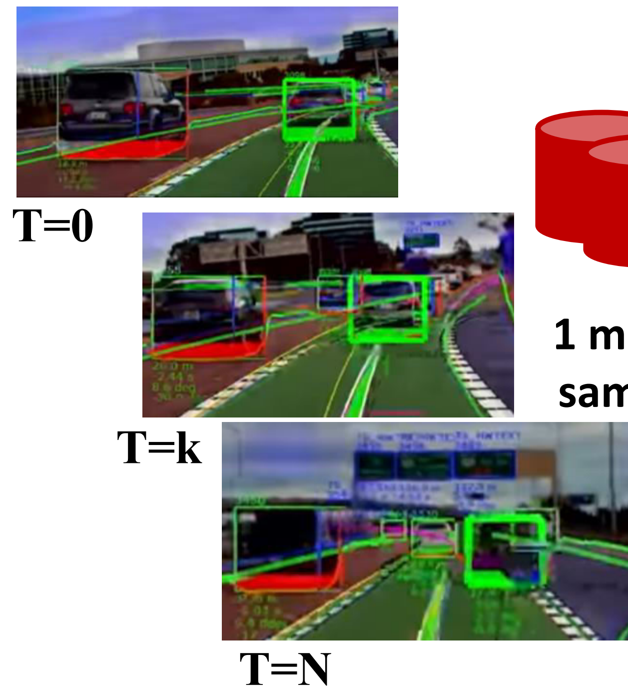
ECE 498 DS : Examples from Real-world

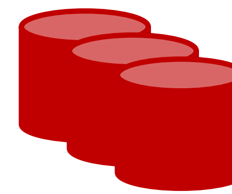
- Uncertainty in measurements, labels etc.
- Structure/Process
- Dataset size: #samples, signal to noise ratio





1000 samples





1 million samples

Seixas et al., *A Bayesian network decision model for supporting the diagnosis of dementia, Alzheimer's disease and mild cognitive impairment*

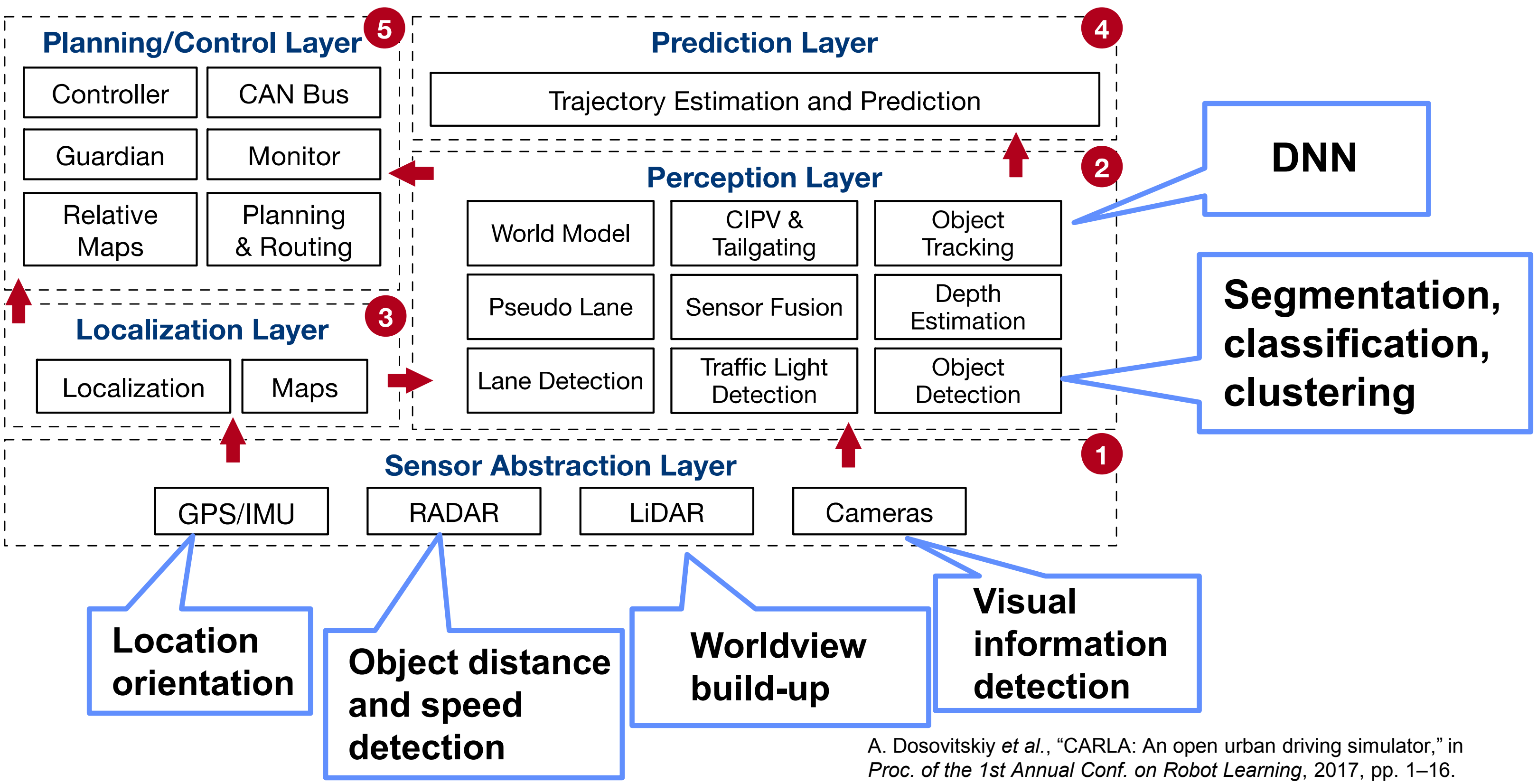
Does the patient suffer from Dementia ?

Identify object in front of the AV?

Integrating AI Techniques in a Real System: Autonomous Vehicles (AVs)

Please follow the link below to the video introducing AV Technology by Waymo
Source: https://www.youtube.com/watch?v=B8R148hFxPw

Autonomous Driving System (ADS)



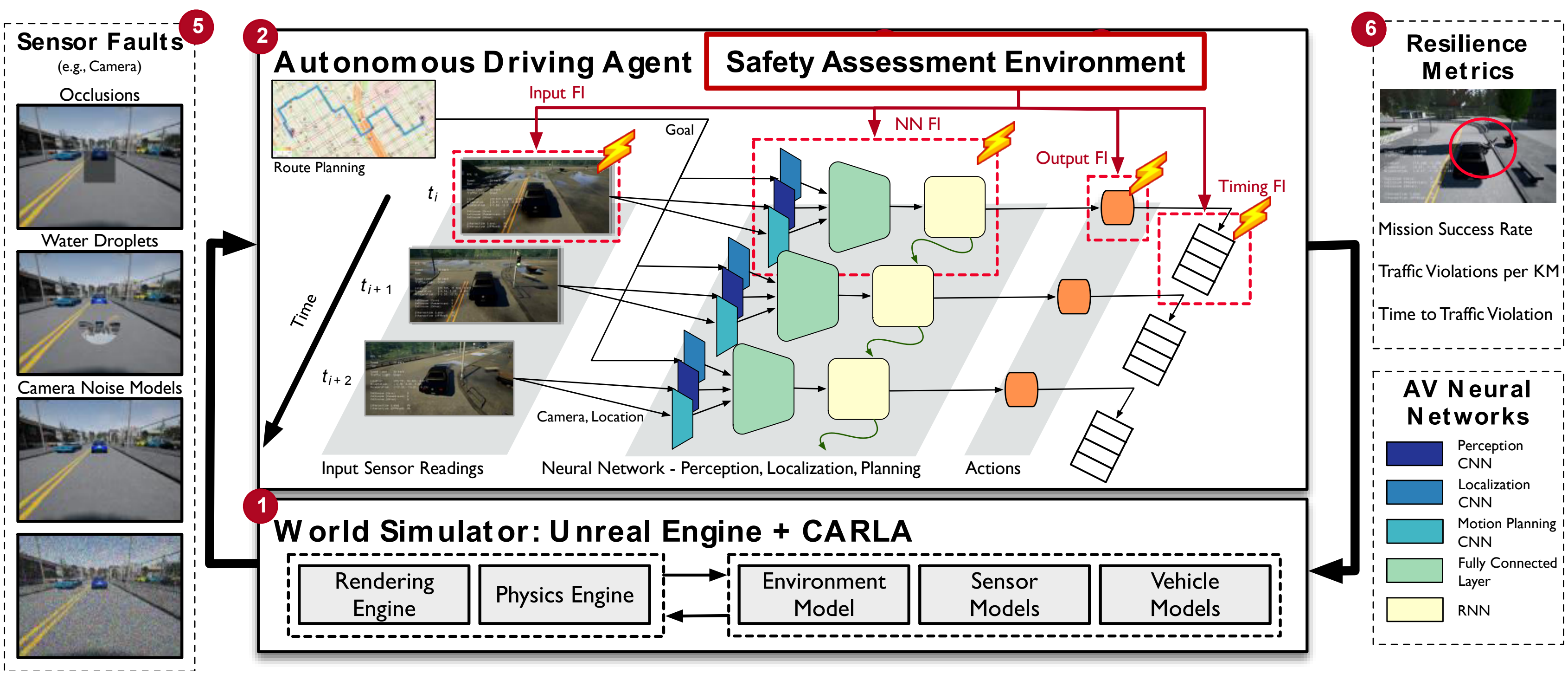
A. Dosovitskiy *et al.*, "CARLA: An open urban driving simulator," in *Proc. of the 1st Annual Conf. on Robot Learning*, 2017, pp. 1–16.

Baidu, "Apollo Open Platform," http://apollo.auto, Accessed: 2018-09-02.

Hands-off the Wheel?

Please follow the link below to the video showcasing AV problems
Source: https://www.youtube.com/watch?v=spw176TZ7-8

AV Safety, Reliability, and Dependability Analytics



Intro to Autonomous Vehicles (AVs)

- Autonomous vehicles (AVs) are complex systems that use artificial intelligence (AI) and machine learning (ML) to integrate mechanical, electronic, and computing technologies to make real-time driving decisions.

- AI enables AVs to navigate through complex environments while maintaining a safety envelope that is continuously measured and quantified by onboard sensors (e.g., camera, LiDAR, RADAR).

- Clearly, the safety, resilience, and dependability of AVs are of a significant concern.

Mini Project 1: AV Safety, Reliability, and Dependability Analytics

- Recent media attention on Tesla/Waymo/Uber AVs
- Resilience and Safety characteristics vary across computing kernels and computing systems
- **Methods and tools for Assessing End-to-End Resilience of AV Technology is not available**

TRANSPORTATION | UBER | RIDE-SHARING

Uber self-driving car saw pedestrian but didn't brake before fatal crash, feds say

The report is more interesting for what it doesn't say than what it does

By Andrew J. Hawkins | @andyjayhawk | May 24, 2018, 11:07am EDT

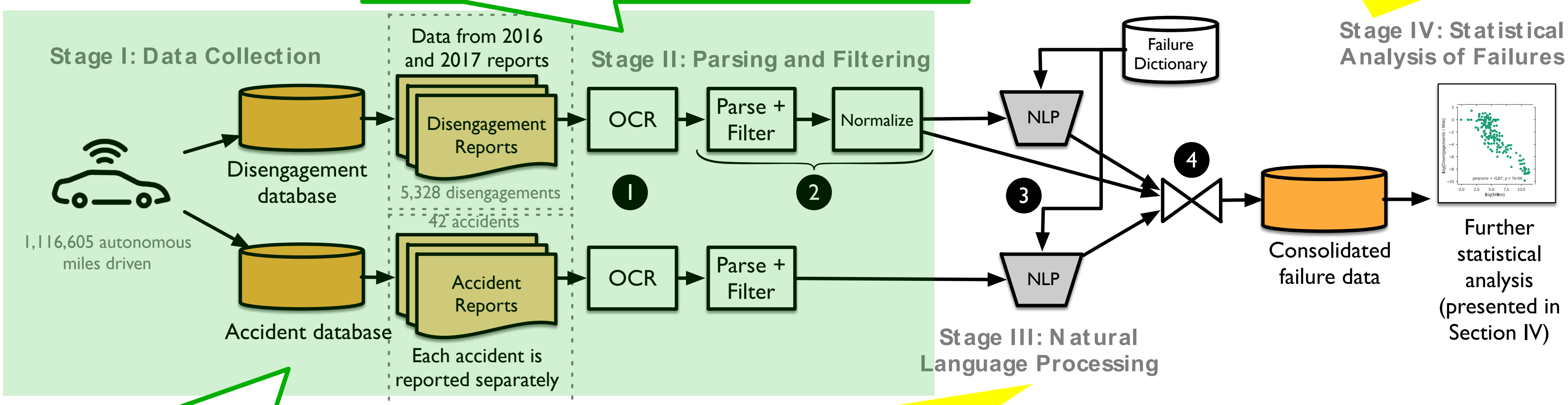
Safety and Reliability Issues [Banerjee et al., DSN 2018]

- **Data and Machine Learning:** 64% of reports were the result of problems in, or untimely decisions made by, the machine learning system
- **Compute system-related**: 30% or more due to failures in computing stack
- **System Design and Integration:** Mismatch between real-world driving behavior, machine learning methods, automotive engineering and compute systems

An End-to-End Workflow for AV Log Data Analysis

- Reports stored as scanned documents.
- Vendor specific parsing & filtering.
- Standardizing data formats across vendors.

- Analyze failure data to quantify
  - Causes
  - Dynamics
  - Impacts



- Vendors are required to collect data as per CA laws.
- CA DMV curates databases of vendor reports.
- No standardized reporting formats.

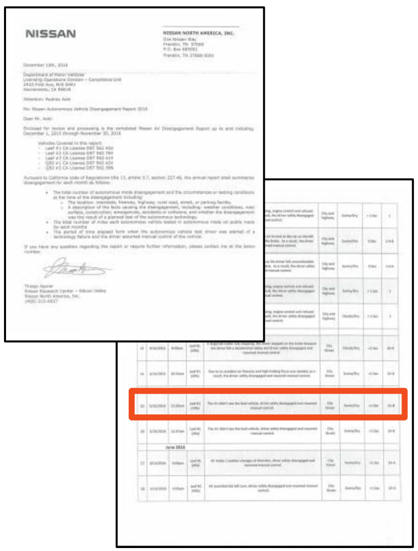
- Localize failures in abstract system model.

Nissan Case Study

2 Individual Report

15	5/25/2016	11:20am	Leaf #1 (Alfa)	The AV didn't see the lead vehicle, driver safely disengaged and resumed manual control.	City Street	Sunny/Dry	<1 Sec	2A-B

3 OCR + Parsing + Cleaning



The AV **didn't see** the lead vehicle…

Analysis shows that AVs do worse than human drivers

Categories: Recognition

1 Nissan Disengagement Reports from the CA DMV

AV Dependability in Adversarial ML

[1]

original fog original rain original translation(40,40) original scale(2.5x)

Weather Conditions

original shear(0.1) original rotation(6 degree) original contrast(1.8) original brightness(50)

Camera and Lighting Conditions

Effect of fault injection

[1] Eykholt, Kevin, et al. "Robust Physical-World Attacks on Deep Learning Visual Classification." *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2018.

[2] Jha, Saurabh, et al. "AVFI: Fault injection for autonomous vehicles." *2018 48th Annual IEEE/IFIP International Conference on Dependable Systems and Networks Workshops (DSN-W)*. IEEE, 2018.

Homework 0

- Questions on application of probability in the computing
- Questions cover
  - Basic Probability
  - Binomial/Poisson Distributions
  - Exponential Distribution
  - Joint Distribution
  - Variance/Mean/Expectation
  - Law of large numbers/Central Limit Theorem
- Be concise and provide all steps
- **Release Date : Jan 16, 2019**
- **Due Date: Jan 23, 2019, 23:59 hrs**
- Students should submit their solutions on Compass (Students are encouraged to type their solutions).

Course Outline I

Week 1	1/14	Course Outline and Overview of Mini Projects a) Autonomous Vehicles (AVs) Safety, Reliability, and Dependability Analytics – California DMV database example b) Healthcare analytics – Genomics/Cancer Example c) Security analytics – Equifax Attack Example Overview of key data analytics and ML concepts
	1/16	Probability Basics Overview
Week 2	1/21	Martin Luther King, Jr. Day – No Lecture
	1/23	In-class Lab 1: Brief tutorial on Python Jupyter Notebook Introductory in-class lab on mini-project 1: Descriptions and Task Specifications Release of mini-project 1
Week 3	1/28	Probability and Hypothesis Testing, p-value; fitting distributions (KS test, KL divergence)
	1/30	In-class Activity 1 (tentative)
Week 4	2/4	Conditional probability, conditional independence;
	2/6	Naïve Bayes, and Bayesian Networks (Continued)
Week 5	2/11	Feature Engineering and Dimensionality Reduction using PCA
	2/13	Unsupervised techniques: statistical and hierarchical clustering
	2/17	Mini-project 1 Due
Week 6	2/18	Unsupervised techniques: Expectation Maximization (EM); GMM and other mixture models In-class Activity 2 (tentative) on Bayesian Networks and clustering
	2/20	Preparation for Mini-project 2: Introduction to Health-care Domain: Disease Models, Drug Response, Forecasting Disease Progression Release of Mini-project 2 Implementing end-to-end workflow for understanding breast-cancer-causing genetic factors Mini-project 1 Presentations (out of class)
Week 7	2/25	Guest Lecture by a Mayo clinician from Center for Individualized Medicine
	2/27	In-class Lab 2 on Mini-project 2
Week 8	3/4	Introduction to Probabilistic Graphical Models; revisit Bayesian Networks
	3/6	In-class Activity 3 (tentative) for Midterm revision

Course Outline II

Week 9	3/11	Midterm
	3/13	Midterm Discussion Markov Models and Hidden Markov Models (HMMs)
Week 10	3/18	Spring Break
	3/20	
Week 11	3/25	Introduction to Factor Graphs
	3/27	Autonomous Security Monitoring for Enterprise Systems: Data-driven learning and inference Mini-project 3 Release: Implementing end-to-end workflow for application of Factor Graphs in security Build a Factor Graph that captures the progression of a realistic multi-stage attack, i.e., the Equifax data breach (2017).
Week 12	4/1	Factor Graphs Continued
	4/3	Pair HMMs, In-class Activity 4 (tentative) on HMMs
Week 13	4/8	In-class Lab 3 on Mini-project 3
	4/10	Factor Graph Inference Algorithms (Belief Propagation, Viterbi)
Week 14	4/15	In-class Activity 5 (tentative) on Factor Graphs
	4/17	Supervised Learning (Linear Regression, SVM, etc.)
Week 15	4/22	Perceptron Model and Neural Networks
	4/24	In-class Activity 6 (tentative) on Neural Networks
Week 16	4/29	Intro to Deep Learning Challenges in Deep Learning Compare Deep Neural Nets (DNN) vs. Probabilistic Graphical Models (PGMs) using an Example Dataset
	5/2	Reading Day
Week 17	5/3-5/10	Final Examination Period As per the Exam Calendar