Course Description

Team Name: Two Engineers
Jiayi Li, jl4924, jl4924@columbia.edu
Chi Zhang, cz2465, cz2465@columbia.edu

(All these course offering information is extracted from Vergil, the CS department website, the EE department website and the course website of each course.)

A.1 Data Analytics

EECS6895: ***Topics in Information Processing. TPC: Advanced Big Data Analytics***

The Big Data Analytics area evolves in a speed that was seldom seen in the history. New Software and Hardware tools are emerging and disruptive. Furthermore, its boundary with Artificial Intelligence becomes blurring. In this Advanced Big Data Analytics course, we will devote to something new -- "How far could we achieve to build a brain that mimics human functions through the state-of-the-art computer science and electrical engineering technologies?" What we would like to discuss is not machines that play Games (Chess, Question & Answering quiz, or Go) or recognize voice and face, but how machines could possibly achieve what are unique to the human beings. Our brains can reason, can associate, and can memorize. We have feeling, emotions, ethics and morality, arts, and consciousness.

EECS6690: ***Topics in Data-driven Analysis and Computation. TPC: Statistical Learning in Biological & Information Systems***

Selected advanced topics in data-driven analysis and computation. Content varies from year to year, and different topics rotate through the course numbers 6690 to 6699. Current topic for 6690: Statistical Learning in Biological & Information Systems.

EECS6893: ***Topics in Information Processing: Big Data Analytics***

This course shall provide the fundamental knowledge to equip students being able to handle those challenges. This discipline inherently involves many fields. Because of its importance and broad impact, new software and hardware tools and algorithms are quickly emerging. A data scientist needs to keep up with this ever changing trends to be able to create a state-of-the-art solution for real-world challenges. This Big Data Analytics course shall first introduce the overview applications, market trend, and the things to learn. Then, I will introduce the fundamental platforms, such as Hadoop, Spark, and other tools, e.g., Linked Big Data. Afterwards, the course will introduce several data storage methods and how to upload, distribute, and process them. This shall include HDFS, HBase, KV stores, document database, and graph database. The course will go on to introduce different ways of handling analytics algorithms on different platforms. Afterwards, the course will zoom in to discuss large-scale machine learning methods that are foundations for artificial intelligence and cognitive networks. The course will discuss several methods to optimize the analytics based on different hardware platforms, such as Intel & Power chips, GPU, FPGA, etc. The lectures will conclude with introduction of the future challenges of Big Data, especially on the onging Linked Big Data issues which involves graphs, graphical models, spatio-temporal analysis, cognitive analytics, etc.

COMS4721: ***Machine Learning for Data Science***

A graduate-level introduction to machine learning. The course covers basic statistical principles of supervised machine learning, as well as some common algorithmic paradigms. Additional topics, such as probabilistic modeling and online decision-making, may be covered if time permits.

COMS4995-004: ***Causal Inference for Data Science***

Context: causality in data science; causality and ML; computational social science; Intro to Counterfactuals and Potential Outcomes; The Pearlian Framework; Confounding; Berkson's Paradox; simpson's paradox; Identifying variables for conditioning: back-door criterion; matching effect estimators; Regression estimators of causal effects; the data processing inequality; Conditioning to reduce entropy; Conditioning to remove bias; Self-selection; Instrumental variables; recommender system example; Mechanisms of action and the Front-door criterion; Regression Discontinuity Design; AB testing and interventions; Multifactorial design; Intent to Treat analysis; an (observational) statistical criterion for causation.

COMS6998-010: ***Cloud Computing and Big Data***

This is a seminar course on Cloud Computing and Data Center Networking intended to provide students exposure to recent advances in these areas. The target audience is graduate level students with the knowledge of operating systems, computer networks and web services. It is expected that students will read and present research papers and work on a course project. The course project is intended to motivate students to identify and work on engineering as well as research issues in these areas. The first part of the course will focus on broad range of topics in Cloud Computing - starting with overview of technologies that enable IaaS Cloud such as Amazon EC2, provisioning and monitoring, storage cloud, elasticity and resource provisioning, and cloud applications. The second part of the course will focus on the recent advancements on data center networking and OpenFlow based network architecture design.

COMS4121: ***Computer System for Data Science***

An introduction to large-scale distributed systems with an emphasis on big-data processing and storage infrastructures. Topics include fundamental tradeoffs in distributed systems, techniques for exploiting parallelism, big-data computation and storage models, design and implementation of various well-known distributed systems infrastructures, and concrete exposure to programming big-data applications on top popular, open-source infrastructures for data processing and storage systems. The course is co-taught by Roxana Geambasu (Assistant Prof. in CS), Sambit Sahu (IBM TJ Watson researcher and CS affiliated faculty), and Eugene Wu (Assistant Prof. in CS). Geambasu will teach fundamental concepts of distributed systems, along with the tradeoffs that arise. Sahu will teach various distributed computation models, along with concrete examples of open-source big-data technologies and how they can be programmed. Wu will teach concepts of data modeling, storage, and visualization, along with the tradeoffs they raise.

A.2 Machine Learning

COMS4721: ***Machine Learning for Data Science***

This course provides an introduction to supervised and unsupervised techniques for machine learning. We will cover both probabilistic and non-probabilistic approaches to machine learning. Focus will be on classification and regression models, clustering methods, matrix factorization and sequential models. Methods covered in class include linear and logistic regression, support vector machines, boosting, K-means clustering, mixture models, expectation-maximization algorithm, hidden Markov models, among others. We will cover algorithmic techniques for optimization, such as gradient and coordinate descent methods, as the need arises.

COMS4995-003: ***Applied Machine Learning***

This class offers a hands-on approach to machine learning and data science. The class discusses the application of machine learning methods like SVMs, Random Forests, Gradient Boosting and neural networks on real world dataset, including data preparation, model selection and evaluation. This class complements COMS W4721 in that it relies entirely on available open source implementations in scikit-learn and tensor flow for all implementations. Apart from applying models, we will also discuss software development tools and practices relevant to productionizing machine learning models.

COMS4771: ***Machine Learning***

This course introduces topics in machine learning for both generative and discriminative estimation. Material will include least squares methods, Gaussian distributions, linear classification, linear regression, maximum likelihood, exponential family distributions, Bayesian networks, Bayesian inference, mixture models, the EM algorithm, graphical models, hidden Markov models, support vector machines, and kernel methods. Students are expected to implement several algorithms in Matlab and have some background in linear algebra and statistics.

EECS6720: ***Bayesian models for machine learning***

This intermediate-level machine learning course will focus on Bayesian approaches to machine learning. Topics will include mixed-membership models, latent factor models and Bayesian nonparametric methods. We will also focus on mean-field variational Bayesian inference, an optimization-based approach to approximate posterior learning. Applications of these methods include image processing, topic modeling, collaborative filtering and recommendation systems. We will discuss a selection of these in class.

ELEN4903: ***Topics in Electrical and Computer Engineering. TPC: Machine Learning***

This course provides an introduction to supervised and unsupervised techniques for machine learning. We will cover both probabilistic and non-probabilistic approaches to machine learning. Focus will be on classification and regression models, clustering methods, matrix factorization and sequential models. Methods covered in class include linear and logistic regression, support vector machines, boosting, K-means clustering, mixture models, expectation-maximization algorithm, hidden Markov models, among others. We will cover algorithmic techniques for optimization, such as gradient and coordinate descent methods, as the need arises. This class is part of the Topics in Electrical & Computer Engineering series.

A.3 Deep Learning

ECBM6040: ***Neural Networks & Deep Learning***

Developing features and internal representations of the world, artificial neural networks, classifying handwritten digits with logistics regression, multilayer perceptron, convolutional neural networks, autoencoders and denoising autoencoders, recurrent neural networks, restricted Boltzmann machines, deep belief networks, deep learning in speech and object recognition.

ECBM6070: ***Topics in Neuroscience and Deep Learning. TPC: Fruit Fly Brain as Neuro Information Processor***

Modeling the brain of model organisms with an emphasis on the fruit fly. Drosophila connectomics. Detailed description of the fruit fly's early olfactory and vision circuits. Navigation and the central complex. The architecture of the Fruit Fly Brain Observatory. Canonical circuits and parallel programming models of local processing units of the fruit fly brain. Methods of Deep Learning. Projects in Python.

EECS6894: ***Topics in Information Processing. TPC: Deep Learning for Computer Vision, Speech and Language***

This graduate level research class focuses on deep learning techniques for vision, speech and natural language processing problems. It gives an overview of the various deep learning models and techniques, and surveys recent advances in the related fields. This course uses Keras and theano as the primary programminging tool. However, other toolkits including Tensorflow, Torch, Caffe, MxNet, PaddlePaddle are also welcome. GPU programming experiences are preferred although not required. Frequent paper presentations and a heavy programming workload are expected.

COMS6998-003: ***Advanced Topics in Deep Learning***

This is a seminar course in which the students read, present, and discuss research papers on deep learning. The focus will be mostly on applications in computer vision, but topics in natural language processing, language translation, and speech recognition will also be read and discussed. It is expected that students taking the course will have prior experience with deep learning and neural network architectures such as Convolutional Neural Nets (CNNs), Recurrent Neural Networks (RNNs), Long Short Term Memories (LSTMs), Gated Recurrent Units (GRUs), Variational Autoencoders

(VAEs), Generative Adversarial Networks (GANs), etc. There will be no assignments other than reading and presentations. However, there will be a final project that students will work on throughout the duration of the semester. Enrollment is capped at 25 students. Instructor permission is required to register.

COMS4995-002: ***Deep Learning***

The course covers the fundamental algorithms and methods, including backpropagation, differentiable programming, optimization, regularization techniques, and information theory behind DNN's. Key deep learning architectures are described in detail including convolution neural networks (CNN, ResNet, DenseNet) with derivation from convolutional sparse coding, recurrent neural networks (LSTM and GRU), variational autoencoders (VAEs), and generative adversarial networks (GANs). Deep reinforcement learning is covered with examples, from basic principles of Markov Decision Processes, value-based and policy-based methods, Deep Q-Networks, policy gradient methods, through Monte Carlo tree search and implementation of AlphaZero and DeepStack with Libratus. Topics from visualization of neural networks, attacks against deep nets, and meta learning are presented. Assignments use the TensorFlow/BayesFlow and PyTorch/Pyro programming frameworks, and a final deep learning project is based on a process, data challenge, or research topic.

A.4 Reinforcement Learning

COMS6998-006: ***Bandits Reinforcement Learning***

This course covers several foundational topics in sequential decision making. We start with multi-armed bandits, and proceed to contextual bandits, a generalization motivated by several real-world applications. We cover the main algorithms, performance guarantees and lower bounds, along with applications in the industry. We also discuss "offline off-policy evaluation": performance assessment techniques for real-world online systems. The remainder of the course concerns the more general problem of reinforcement learning, from small-state MDPs to recent extensions that handle large state spaces. *Throughout, the course will primarily focus on the theoretical foundations, with some discussion of the practical aspects.*

ELEN6885: ***Topics in Signal Processing. TPC: Reinforcement Learning***

This course offers an advanced introduction Markov Decision Processes (MDPs)– a formalization of the problem of optimal sequential decision making under uncertainty–and Reinforcement Learning (RL)–a paradigm for learning from data to make near optimal sequential decisions. The first part of the course will cover foundational material on MDPs. We'll then look at the problem of estimating long run value from data, including popular RL algorithms like temporal difference learning and Q-learning.

A.5 Speech/Text Processing and Recognition, **(NLP)**

EECS6894: ***Topics in Information Processing. TPC: Deep Learning for Computer Vision, Speech and Language***

This graduate level research class focuses on deep learning techniques for vision, speech and natural language processing problems. It gives an overview of the various deep learning models and techniques, and surveys recent advances in the related fields. This course uses Keras and theano as the primary programminging tool. However, other toolkits including Tensorflow, Torch, Caffe, MxNet, PaddlePaddle are also welcome. GPU programming experiences are preferred although not required. Frequent paper presentations and a heavy programming workload are expected.

COMS6998-007: ***Fund Speech Recognition***

The first portion of the course will cover fundamental topics in speech recognition: signal processing, Gaussian mixture distributions, the Expectation-Maximization algorithm, deep neural networks, hidden Markov models, pronunciation modeling, decision trees, language modeling, finite-state transducers, and search. Topics will be covered in sufficient detail for students to be able to implement a basic large vocabulary speech recognizer.In the remainder of the course, selected topics from the current state of the art will be discussed. We will cover several key areas in more depth and survey some advanced topics, including acoustic adaptation, discriminative training, and maximum entropy models.

COMS6998-012: ***NLP in Context Computational M***

Computational Models of Social Meaning is a seminar in Natural Language Processing, focusing on computational methods for extracting social and interactional meaning, mainly from text (both traditional media and social media) and speech. Topics include detection of speaker's sentiments, emotions, opinions and beliefs, sarcasm, deception, persuation, perspective, power and influence, politeness, and personality. Analysis of meaning-bearing characteristics of the speaker and topic, including text, discourse, prosodic and other cues.

A.6 Computer Vision/ Image Processing

EECS6894: ***Topics in Information Processing. TPC: Deep Learning for Computer Vision*, *Speech and Language***

This graduate level research class focuses on deep learning techniques for vision, speech and natural language processing problems. It gives an overview of the various deep learning models and techniques, and surveys recent advances in the related fields. This course uses Keras and theano as the primary programminging tool. However, other toolkits including Tensorflow, Torch, Caffe, MxNet, PaddlePaddle are also welcome. GPU programming experiences are preferred although not required. Frequent paper presentations and a heavy programming workload are expected.

ELEN4830: ***Digital Image Processing***

This course will introduce fundamental technologies for digital image/video representation, compression, analysis, and processing. Students will gain understanding of algorithm/system design, analytical tools, and practical implementations of various digital image applications. During this semester, we will update the course content to increase hands on experience in applying analytical solutions in practical applications. Topics include digital image/video perception, sampling, optimal quantization, halftoning, transform, filtering, multi-spectral processing, restoration, analysis, feature extraction, morphological transform, coding, segmentation, and 3D model reconstruction. Considerations of practical system requirements (e.g.medical, satellite, consumer) will be discussed as well.

COMS4731: ***Computer Vision***

This course provides an introduction to Computer Vision and is intended for advanced undergraduate and graduate students (MS and/or Ph.D.). We require a mathematical background (e.g., calculus, linear algebra, some knowledge of probability and statistics) and the ability to program in MATLAB. This will be a very *hands-on* course where we will both learn AND program the concepts we are learning. Computer Vision is as much about application as it is about theory. If you are unsure about your background, please email me or come see me. We cover the following topics (listed in more detail below): image formation, image sensing, binary processing, image processing, segmentation, feature detection/matching/tracking, optical flow, image stitching, face detection, computational photography, appearance tracking, camera calibration, pose estimation, stereo vision, structure-from-motion, structured light, 3D reconstruction (shape-from-X, depth-from-defocus), photometric stereo, appearance matching, and recognition.

A.7 Distributed Systems

COMS6998-003: ***Advanced Distributed System***

This class provides an overview of influential research that provided the basis of most large-scale, cloud infrastructures today. Students read, present, and discuss papers on important distributed systems topics, including distributed consensus, consistency models and algorithms, service-oriented architectures, large-scale data storage, and distributed transactions, big-data processing frameworks, and distributed systems security.

COMS4113-001: ***Fundamentals of Large-Scale Distributed Systems***

Core concepts of large-scale distributed systems; The inner-workings of several distributed systems serving as infrastructure for some very big companies; Gothas of using some popular distributed systems, which stem from their inner workings and reflect the challenges of building large-scale distributed systems; How to build a real distributed system yourself!

A.8 Internet of Things

EECS4764: ***Internet of Things - Intelligent and Connected Systems***

Cyber-physical systems and Internet-of-Things. Various sensors and actuators, communication with devices through serial protocols and buses, embedded hardware, wired and wireless networks, embedded platforms such as Arduino and smartphones, web services on end devices and in the cloud, visualization and analytics on sensor data, end-to-end IoT applications. Group projects to create working CPS/IoT system.

EECS6765: ***Internet of Things: Systems and Physical Data Analytics***

Internet of Things from the point of view of data. Methods for data analytics to understand tradeoffs and partitioning between cloud-based data-analytics and physical-device data-analytics. Two-way interaction between data and physical devices to support a truly ubiquitous, networked and autonomous cyber-physical ecosystem. System-focused design of architectures, algorithms, networks, protocols, communications, power, security and standards. Focus on a significant design project.

A.9 Computer Networks

CSEE4119: ***Computer Networks***

The course will cover the core elements of modern Internet technology and protocols, including the application, transport, network, link layers and physical layers, for both wired and wireless networks. Coverage roughly corresponds to Chapters 1-8 of the textbook, with selected materials from the later chapters, and additional instructor-provided resources.

COMS4180: ***Network Security***

Introduction to network security concepts and mechanisms. Foundations of network security and an in-depth review of commonly-used security mechanisms and techniques, security threats and network-based attacks, applications of cryptography, authentication, access control, intrusion detection and response, security protocols (IPsec, SSL, Kerberos), denial of service, viruses and worms, software vulnerabilities, web security, wireless security, and privacy.

CSEE4140: ***Networking Lab***

Networking Laboratory is a course where you actively learn by doing. You will learn network concepts and protocols by configuring a network using routers and PCs, observing the actual behavior of these machines and the overall network, and analyzing and evaluating the results. This is a learning experience different from other networking courses where you read and think in an abstract level. In this course, you need to get real results by getting your hands dirty and taking your time configuring the machines properly. We have a cool (=awesome + cold) laboratory that only enrolled students can enter and use just for that purpose. Topics covered include IP, ARP, ICMP, RIP, OSPF, BGP, TCP, UDP, STP, DNS, NAT, DHCP, SNMP, IGMP and PIM-DM/SM.

ELEN6771: ***Topics in Networking. TPC: 5G Programmable Networks***

Further study of areas such as communication protocols and architectures, flow and congestion control in data networks, performance evaluation in integrated networks. Content varies from year to year, and different topics rotate through the course numbers 6770 to 6779.

ELEN6772: ***Topics in Networking. TPC: Network Algorithms and Learning***

Further study of areas such as communication protocols and architectures, flow and congestion control in data networks, performance evaluation in integrated networks. Content varies from year to year, and different topics rotate through the course numbers 6770 to 6779.

ELEN6770: ***Topics in Networking. TPC: Next Generation Networks***

Further study of areas such as communication protocols and architectures, flow and congestion control in data networks, performance evaluation in integrated networks. Content varies from year to year, and different topics rotate through the course numbers 6770 to 6779.

ELEN6775: ***Topics in Networking. TPC: Advanced Networks: A Systems Approach***

Further study of areas such as communication protocols and architectures, flow and congestion control in data networks, performance evaluation in integrated networks. Content varies from year to year, and different topics rotate through the course numbers 6770 to 6779.

ELEN6776: ***Topics in Networking. TPC: Content Distribution Networks***

Further study of areas such as communication protocols and architectures, flow and congestion control in data networks, performance evaluation in integrated networks. Content varies from year to year, and different topics rotate through the course numbers 6770 to 6779.

A.10 Neuroscience

EEBM6091: ***Topics in Computational Neuroscience and Neuro engineering. TPC: Devices and Analysis for Neural Circuits***

Devices and Analysis for Neural Circuits.

BMEB4020: ***Computational Neuroscience: Circuits in the Brain***

The biophysics of computation: modeling biological neurons, the Hodgkin-Huxley neuron, modeling channel conductances and synapses as memristive systems, bursting neurons and central pattern generators, I/O equivalence and spiking neuron models. Information representation and neural encoding: stimulus representation with time encoding machines, the geometry of time encoding, encoding with neural circuits with feedback, population time encoding machines. Dendritic computation: elements of spike processing and neural computation, synaptic plasticity and learning algorithms, unsupervised learning and spike time-dependent plasticity, basic dendritic integration. Projects in MATLAB.

BMEE4030: ***Neural Control Engineering***

Topics include: basic cell biophysics, active conductance and the Hodgkin-Huxley model, simple neuron models, ion channel models and synaptic models, statistical models of spike generation, Wilson-Cowan model of cortex, large-scale electrophysiological recording methods, sensorimotor integration and optimal state estimation, operant conditioning of neural activity, nonlinear modeling of neural systems, sensory systems: visual pathway and somatosensory pathway, neural encoding model: spike triggered average (STA) and spike triggered covariance (STC) analysis, neuronal response to electrical micro-stimulation, DBS for Parkinson's disease treatment, motor neural prostheses, and sensory neural prostheses.

ECBM4040: ***Neural Networks and Deep Learning***

Developing features & internal representations of the world, artificial neural networks, classifying handwritten digits with logistics regression, feedforward deep networks, back propagation in multilayer perceptrons, regularization of deep or distributed models, optimization for training deep models, convolutional neural networks, recurrent and recursive neural networks, deep learning in speech and object recognition.

ECBM4060: ***Introduction Genomic Information Science & Technology*** **ECBM4090:** ***Brain Computer Interfaces Lab***

Introduction to the information system paradigm of molecular biology. Representation, organization, structure, function, and manipulation of the biomolecular sequences of nucleic acids and proteins. The role of enzymes and gene regulatory elements in natural biological functions as well as in biotechnology and genetic engineering. Recombination and other macromolecular processes viewed as mathematical operations with simulation and visualization using simple computer programming.

COMS6998-007: ***Introduction Brain Computer Interact***

Introduction Brain Computer Interact

A.11 Security

COMS6998-009: ***Security and Robustness of ML Systems***

Over the past few years, Machine Learning (DL) has made tremendous progress, achieving or surpassing human-level performance for a diverse set of tasks including image classification, speech recognition, and playing games such as Go. These advances have led to widespread adoption and deployment of ML in security- and safety-critical systems such as self-driving cars, malware detection, and aircraft collision avoidance systems. This wide adoption of DL techniques presents new challenges as the predictability and correctness of such systems are of crucial importance. Unfortunately, ML systems, despite their impressive capabilities, often demonstrate unexpected or incorrect behaviors in corner cases for several reasons such as biased training data, overfitting, and underfitting of the models. In safety- and security-critical settings, such incorrect behaviors can lead to disastrous consequences such as a fatal collision of a self-driving car. For example, a Google self-driving car recently crashed into a bus because it expected the bus to yield under a set of rare conditions but the bus did not. A Tesla car in autopilot crashed into a trailer because the autopilot system failed to recognize the trailer as an obstacle due to its "white color against a brightly lit sky" and the "high ride height." Such corner cases were not part of Google's or Tesla's test set and thus never showed up during testing. Other examples include Microsoft's Tay chatbot tweeting racist words because it was misled by malicious twitter users, and Google removing "gorilla" as an image class after its image classification algorithm incorrectly classified dark skined people as gorillas.

COMS 4180: ***Network Security***

Introduction to network security concepts and mechanisms. Foundations of network security and an in-depth review of commonly-used security mechanisms and techniques, security threats and network-based attacks, applications of cryptography, authentication, access control, intrusion detection and response, security protocols (IPsec, SSL, Kerberos), denial of service, viruses and worms, software vulnerabilities, web security, wireless security, and privacy.

COMS 4187: ***Security Architecture and Engineering***

Computer security concepts and mechanisms; measures employed in countering such threats. Concepts and tools available in order to assume an appropriate security posture. Foundations of security. Identification, authentication, authorization. Software design for security and assurance. Hardware assists. Security architecture; design for security. Security tradeoffs.