Jeffrey Lee

Education

Undergraduate **Carnegie Mellon University**, *B.S. in Computer Science, Minor in Machine Learning*, 2016-2020, GPA 3.21.

Graduate **University of Oklahoma**, *Master's in Meteorology*, 2021-present, GPA 4.00.

Selected Coursework

2023 Advanced Synoptic Meteorology - An in-depth study of the synoptic environment. Topics covered include geostrophic theory, omega equation and height-tendency equation, Trenberth formulation, Q vectors, QG potential vorticity, and Isentropic potential vorticity.

2022 Atmospheric Radiation - An introduction into the principles and practical consequences of atmospheric radiation as needed to understand the role of atmospheric radiation in meteorological and climate studies. Topics to be discussed include the theory of radiative transfer, electromagnetic waves, the electromagnetic spectrum, spectra of gaseous molecules, reflection and refraction, radiative properties of objects, thermal emission, atmospheric transmission, atmospheric emission and absorption, broadband fluxes and heating rates, and scattering and absorption by aerosols. To apply the theory, numerical computation of atmosphere radiation and light scattering will also be covered in this course.

2022 Advanced Statistical Meteorology - An overview of some advanced statistical methods used to interpret data in the atmospheric and oceanic sciences. Major topics include regression/correlation and epoch analyses, time series analysis (power spectra, filtering), matrix methods for signal decomposition (EOFs, CCA), and objective mapping and covariance modeling.

2021 Fundamentals of Atmospheric Science - A rigorous survey of the fundamental concepts in atmospheric science to provide the foundation for future graduate course work in meteorology and in related disciplines. Topics include atmospheric dynamics, midlatitude meteorology, atmospheric thermodynamics, atmospheric radiation, cloud microphysics, atmospheric chemistry, planetary boundary layer, climate dynamics, radar meteorology, tropical meteorology, numerical weather prediciton, data assimilation, and polar meteorology.

2020 Advanced Deep Learning - Models that are capable of extracting complex, hierarchical representations from high-dimensional data lie at the core of solving many ML and AI domains, such as visual object recognition, information retrieval, natural language processing, and speech perception. While the usefulness of such deep learning techniques is undisputed, our understanding of them is still in many ways nascent. The goal of this course is to introduce students to recent and exciting developments (both theoretical and practical) in these methods. List of topics include representational power of neural networks, generalization in neural networks, optimization for deep learning, graphical models, latent-variable models, variational inference, MCMC, autoregressive models, autoencoders, GANs, self-supervised/predictive learning, word embeddings, RNNs, Seq2Seq architectures, and attention models.

2019 Convex Optimization - Many problems of interest in machine learning can be posed as optimization tasks that have special properties — such as convexity, smoothness, sparsity, separability, etc. — permitting standardized, efficient solution techniques. This course is designed to give a graduate-level student a thorough grounding in these properties and their role in optimization, and a broad comprehension of algorithms tailored to exploit such properties.

2019 Deep Reinforcement Learning - This course brings together many disciplines of Artificial Intelligence (including computer vision, robot control, reinforcement learning, language understanding) to show how to develop intelligent agents that can learn to sense the world and learn to act by imitating others, maximizing sparse rewards, and/or satisfying their curiosity. Topics covered include imitation learning, Monte Carlo learning, policy gradient methods, actor critic, evolutionary methods, Sim2Real, and model based reinforcement learning.

2018 Modern Regression - This course is an introduction to the real world of statistics and data analysis. Topics covered include simple linear regression, multiple linear regression, variable selection, interaction terms, transformation of variables, resampling and bootstrap for statistical inference, and regression trees.

Projects and Experience

PurpleAir PM Validation (2023) Created a framework for comparing low-cost PurpleAir air quality sensors against federally regulated monitors. Used to extrapolate PM distributions from PurpleAir sensors as input to the smoke module of the Rapid Refresh Forecast System being developed at NOAA. Led by Hongli Wang.

ML for Surface $PM_{2.5}$ Predictions (2021 - present) Smoke aerosols from wildfires play a large part in determining local weather phenomena and affect human health. However, spatial coverage of reliable air quality data is limited to a sparse network of air quality monitoring stations. To bridge this gap in data, I am creating an ML model to predict surface $PM_{2.5}$ concentrations using satellite data and ground-based observations. Led by Dr. Marcela Loría-Salazar.

ML for Veil Cloud Identification (2019 - 2020) Veil clouds cover a significant percentage of the Earth's oceans, but their formation and interaction with man-made aerosols is not well understood. This project focused on applying deep learning to atmospheric data collected from NOAA and NASA satellites to accurately identify and model veil clouds. Led by Dr. Hamish Gordon.

Delta Air Lines Internship (2019) Built a web application for centralizing and tracking internal IT requests. Learned several frameworks under the Delta full stack. Familiarized myself with modern IT architectural designs such as CICD pipelines.

DDQN (2019) Used double q-learning to solve `Cartpole-v0` and `MountainCar-v0` environments in Gym by OpenAI. Included a hindsight experience replay (HER) to combat maximization bias.

Dementia Prediction (2018) Created an adaptive model built to perform classification of patient cognition status over multiple timepoints. This model was created for the NACC dataset. I employed several layers of statistical and machine learning models to not only provide the user with both short- and long-term predictions, but also a predictor ranking that would inform clinicians which tests would be most efficient in terms of information gain. Developed under Dr. Chi at the University of Minnesota.

Technical Skills

Languages Fluent in Java, Python, C, and R

Toolset Comfortable working in both Windows and Linux. Experienced with IDE's, version control with git, AWS, Tensorflow, PyTorch, NASA satellite aerosol retrievals.