

ENGR 5963 Machine Learning Engineering for Production (MLOps)

5000 level 2 Semester Hour Course

Course Description

While many data scientists regularly build machine learning models in notebooks many have never put those models into production. The successful deployment of machine learning models requires additional knowledge and skills related to the continual deployment and monitoring of machine learning models in a constantly changing world. This course expects students to have knowledge of building machine learning models using notebooks and will learn how to deploy those models in production. Class sessions are broken into two parts, the first half covering theory and the second half a hands-on lab using python and other open-source tools.

Learning Objectives

Learning objectives for the course are:

- Understanding the ML Lifecycle and Deployment
- Cleaning, and validating datasets and assessing data quality
- Data definitions, data drift and concept drift
- Automated feature selection
- Feature engineering
- Hyperparameter optimization
- Assessing model quality
- Explainable AI and model interpretability
- Data storage and data lakes
- Advanced Data Labeling Methods
- Data Augmentation
- Data dependent preprocessing
- Neural Architecture Search
- Resource management
- Deploying Machine Learning Models in Production
- Model Monitoring and Logging
- Dynamic model deployment

Course Prerequisites

Programming experience with python. An intro to machine learning course is required.

Qualifications

PhD in Computer Science from UCLA with a minor field in statistics
MS in Information Design and Visualization from Northeastern University
All of my consulting work involved developing production machine learning systems for various companies.

Preferences

I prefer the classes meet in 3 hour sessions in the afternoons or late mornings.
Giving students the option of on-ground or remote through NuFlex makes the most sense to me.