CE 784: Machine Learning and Data Analytics for Civil Engineering Applications

Syllabus

Objectives: The objective of this course is to provide the students with the introduction of machine learning (ML) and large-scale data analytics tools with their applications in civil engineering. The course will emphasize on 1) traditional supervised algorithms such as support vector machines, 2) traditional unsupervised machine learning algorithms such as k-means clustering, 3) deep learning algorithms such as convolution neural networks, 4) fundamentals of tools used to handle large-scale data such as map-reduce, and 5) visualizing large scale data-bases. Fundamentals of these algorithms and tools and their applications in different real-world problems related to civil engineering will be covered along with a course project.

Pre-Requisites: Basic Probability and Statistics, Linear Algebra, Basic Programming Skills in Python

Course Contents:

No.	Broad Title	Topics
1.	Introduction and Background	Introduction, Historical context, Necessities, ML in modern civil engineering, Real-world application examples. Recapitulation of linear regression. Logistic regression
2.	Shallow Supervised Algorithms	K-Nearest Neighbor, Neural Networks Learning, Backpropagation, Support Vector Machines, Applications to structural damage detection, soil classification, etc.
3.	Convolutional Neural Networks	Introduction to ConvNets, activation functions, hyper-parameter tuning, dropout, batch normalization. Applications to camera-based classification and object detection related to structural health monitoring, vehicle detection, pavement distress detection, etc.
4.	Unsupervised Clustering	Hierarchical clustering, K-means clustering, Density based clustering. Applications on transportation mode inference, level of service of roads, etc.
5.	Tools and Software	Data Visualization tools such as Tableau, Power BI and Deep learning tools such as pytorch, keras, etc.
6.	Recurrent Neural Networks	Recurrent Neural Networks, Long-Short Term Memory. Applications to traffic state (speed, volume) prediction, soil strength prediction, rainfall-runoff modelling, etc.
7.	Generative Models	Variational Autoencoder, Generative Adversarial Networks. Applications to sensor data generation and imputation such as traffic sensors, fault diagnostics in structural health monitoring, etc.
8.	Hadoop Map-Reduce and Spark	Map-Reduce fundamentals (key-value), interface, algorithms (matrix multiplication, sorting, etc.), Apache Pig, Hive, Spark Fundamentals, RDD, Spark Streaming. Applications to large-scale traffic trajectory data analysis, building information modelling in construction industry, etc.
9.	Project*	Discussion and presentation of course project

* 8-week long course project where students will apply the tools/algorithms covered in the course on a topic of their choice of interest.

Recommended Books:

Textbooks: None

Reference Books:

a. "The Elements of Statistical Learning: Data Mining, Inference and Prediction", *Trevor Hastie, Robert Tibshirani, Jerome Friedman*, Springer
b. "Pattern Recognition and Machine Learning", *Christopher M. Bishop,* Springer
c. "Deep Learning", *Ian Goodfellow, Yoshua Benjio, Aaron Courville*, MIT Press
d. "Hadoop: The Definitive Guide (4th Edition)", *Tom White*, O'Reilly Media

e. “Learning Spark: Lightning Fast Big Data Analysis”, *Matei Zaharia, Holden Karau, Andy Konwinski, Patrick Wendell*, O’Reilly Media
f. “Hadoop in Practice”, *Alex Holmes*, Manning Publications
g. Relevant technical reports, journal, and conference publications
h. Other online resources available