Deep Learning Applications in Statistical Problems

Overview:

This short course delves into the intersection of Deep Learning and statistical analysis. Participants will explore and apply Deep Learning methodologies to tackle various statistical problems. The course covers advanced topics such as longitudinal data analysis, survival analysis, quantile regression, autoencoders, generative models, and handling spatial-temporal data using Deep Learning techniques.

Objectives:

- Understand the principles and fundamentals of Deep Learning as applied to statistical problems.

- Gain insights into Survival Analysis and utilize Deep Learning for predictive modeling in this context.

- Implement quantile regression using Deep Learning architectures for robust estimation.

- Comprehend generative models and their applications in data generation and dimensionality reduction.

- Apply Deep Learning to analyze and model spatial-temporal data effectively.

Course Outline:

Morning Session (8:30 am --- 12:30 pm):

1. **Introduction to Deep Learning** (1.5 hrs)

- Brief overview of the development and applications of deep learning (image language decision)

- Introduce some modern neural network structures (MLP CNN Transformer)

- Deep learning theory

2. **Introduction to Deep Learning Packages/resources** (1 hr) runpeng

- Overview of Pytorch, Huggingface, etc.

- Basics: Training models, visualization

- Examples: mnist

3. **Deep Generative Models** (1.5 hrs) Wang

- The relationship of latent variable models, autoencoders, and generative models

- Well-known generative models: VAE, GAN, Flow, and Diffusion model

- Image generation and reconstruction

- mnist

Afternon Session (1:30 pm --- 4:30 pm):

4. **Transformer and self-supervised learning** (1 hr) Wang

5. **Survival Analysis with Deep Learning and Deep Quantile Regression** (1 hr)

- Brief review of some survival analysis questions.

- Utilizing Deep Learning models (such as Cox proportional hazards model) for survival prediction

- Example: Use deep cox model to model the hazard function and predict survival time of breast cancer patient from UK Biobank dataset.

- Brief introduction of quantile regression, comparing to simple regression.

- Introduce the special neural network design for non crossing quantile regression.

- Example: Distributional network for treatment effect measurement.

6. **Time Series and Spatial-Temporal Data Modeling with Deep Learning** (1 hr)

- Introduction to recurrent neural network (RNN GNN LSTM).

- Introduce the task of time series prediction (Temporal) and OD prediction (Spatial-Temporal).

- Example: Employ transformer-based architecture for predicting origin-destination matrices. Deep-set models and offline reinforcement learning framework for treatment effect evaluation in the context of spatiotemporal interference.

Target Audience:

This course is designed for statisticians, data scientists, researchers, and professionals in the field of statistics and data analysis seeking to deepen their understanding of applying Deep Learning to solve complex statistical problems.

Prerequisites:

- Proficiency in statistics and data analysis

- Basic knowledge of machine learning and neural networks

- Familiarity with python and pytorch

Conclusion:

Upon completion of this course, participants will have gained advanced knowledge and practical skills in applying Deep Learning methodologies to address a variety of statistical challenges, enabling them to tackle real-world problems effectively.

Instructor:

Hongtu Zhu:

Dr. Hongtu Zhu is a professor of Biostatistics in the University of North Carolina at Chapel Hill. He has a broad background in statistics, biostatistics, medical imaging, genetics and computational neuroscience, with specific training and expertise in neuroimaging data analysis and big data integration as well as secondary data analysis on neurodegenerative and neuropsychiatric diseases.

Xiao Wang:

Dr. Xiao Wang is a professor of statistics at Purdue University. He received his Ph.D. from the University of Michigan, and his research centers on machine learning, nonparametric statistics, and functional data analysis. Dr. Wang's work has been featured in leading statistical journals and machine learning conferences, and he is a fellow of the Institute of Mathematical Statistics (IMS) and the American Statistical Association (ASA). Dr. Wang is currently an associate editor for JASA, Technometrics, and Lifetime Data Analysis.

Runpeng Dai

Runpeng is a first year PhD student in the University of North Carolina at Chapel Hill. His research interest lies in Reinforcement learning, Learning theory and spatial-temporal data analysis. He has several internship experiences in ride-sharing and Quantitative Finance companies.