Indiana University Indianapolis
Department of Mathematical Sciences

STATISTICS SEMINAR

12:15pm—1:15pm, Tuesday, November 12, 2024
Zoom Meeting: Meeting ID: 845 0989 4694

Speaker: Yue Xing
Department of Statistics and Probability, Michigan State University

Title: Benefits of Transformer Architecture: A Study in In-Context Learning

Abstract:

In practice, it is observed that transformer-based models can learn concepts in context in the inference stage. While existing literature, e.g., Zhang et al. (2023), provide theoretical explanations on this in-context learning ability, they assume the input x_i and the output y_i for each sample are embedded in the same token (i.e., structured data). However, in reality, they are presented in two tokens (i.e., unstructured data in Wibisono et al., 2023). We conduct experiments in linear regression tasks to study the benefits of the architecture of transformers and provide some corresponding theoretical analysis to explain why the transformer can learn from unstructured data. We study the exact components in a transformer that facilitate the in-context learning. In particular, we observe that (1) a transformer with two layers of softmax (self-)attentions with look-ahead attention mask can learn from the prompt if y_i is in the token next to x_i for each example; (2) positional encoding can further improve the performance; and (3) multi-head attention with a high input embedding dimension has a better prediction performance than single-head attention. Finally, for large language models (LLMs) used in real practice, we introduce a tool to facilitate the analysis of their ICL behavior.

Bio:
Dr. Yue Xing is an Assistant Professor at the Department of Statistics and Probability in Michigan State University. She previously obtained her Ph.D. degree in Statistics at Purdue University and her M.Phil and bachelor's degrees at the Chinese University of Hong Kong . Her research interests include trustworthy AI (attack and robustness, privacy, watermark) and the theory of generative AI.