

Transformers serve as the foundational architecture for large language and video generation models, such as GPT, BERT, SORA, and their successors. While empirical studies have shown that real-world data and learning tasks exhibit low-dimensional geometric structures, the theoretical understanding of transformers in leveraging these structures remains largely unexplored. In this talk, we present a theoretical foundation for transformers in two key scenarios: (1) regression tasks with noisy input data lying near a low-dimensional manifold, and (2) in-context learning (ICL) for regression of Hölder functions on manifolds. For the first setting, we prove that approximation and generalization bound that depend crucially on the intrinsic dimension of the manifold, demonstrating that transformers can effectively learn from data perturbed by high-dimensional noise. For the second setting, we derive generalization error bounds for ICL in terms of prompt length and the number of training tasks, revealing that transformers achieve the minimax optimal rate for Hölder regression—scaling exponentially with the intrinsic rather than ambient dimension. Together, these results provide foundational insights into how transformers exploit low-dimensional geometric structures in learning tasks, advancing our theoretical understanding of their remarkable empirical success.