

SEMINARI I DEPARTAMENTIT TË MATEMATIKËS “Math UP Seminar”

Kemi kënaqësinë t'ju njoftojmë se ligjërata e radhës në kuadër të Seminarit të Departamentit të Matematikës në FSHMN ka folës:

Dr. Ismail Elezi

Orari: 3 Tetor 2024, Salla 113, 13:00—14:00

Titulli: Enhancing Neural Network Efficiency: From Knowledge Distillation to Memory-Efficient Training

Abstract

Recent advancements in training large neural networks have produced promising results in both efficiency and performance. We present two novel approaches: a constrained feature distillation method for knowledge transfer and a memory-efficient training algorithm for large language models. Our distillation method, based on orthogonal projections and task-specific normalization, outperforms previous state-of-the-art on ImageNet and shows significant improvements in object detection and image generation. For large language models, we introduce a technique that compresses intermediate activations without performance degradation, utilizing rank-1 sub-token projections during forward passes and coarse reconstruction in backward passes. This approach complements existing parameter-efficient fine-tuning methods and demonstrates competitive performance in both fine-tuning and pre-training scenarios. These innovations offer substantial improvements in model efficiency across various tasks and architectures, opening the way for more accessible and powerful neural network applications.