Kuat Gazizov

Merced, CA, United States
kgazizov@ucmerced.edu linkedin.com/in/kgazizov Personal Website

Education

University of California, Merced **Jan 2022 – present**
PhD in Computer Science *Merced, CA, United States*
Supervisor: Miguel Á. Carreira-Perpiñán

Lomonosov Moscow State University **Sep 2015 – Jun 2019**
Bachelor of Science in Mathematics *Moscow, Russia*

Publications

[**NeurIPS 2024**] M. Á. Carreira-Perpiñán and K. Gazizov: "The tree autoencoder model, with application to hierarchical data visualization". *Advances in Neural Information Processing Systems, 2024*

[**BayLearn 2024**] M. Á. Carreira-Perpiñán and K. Gazizov: "Hierarchical data visualization via PCA trees". *Extended abstract at Bay Area Machine Learning Symposium, 2024*

K. Gazizov and A. Zharmagambetov and M. Á. Carreira-Perpiñán: "A critical comparison of soft vs hard oblique classification trees". (**In submission**)

[**BayLearn 2023**] K. Gazizov and A. Zharmagambetov and M. Á. Carreira-Perpiñán: "Pros and cons of soft vs hard decision trees". *Extended abstract at Bay Area Machine Learning Symposium, 2023*

Projects

Layer-wise Neural Network Quantization | *QAT, Adaptive Mix-bit Quantization, Optimization*

- Developing a neural network compression method by formulating the quantization problem as a constrained optimization problem. The approach involves applying augmented Lagrangian and alternating optimization techniques to determine optimal per-layer bit-widths for quantization. The goal is to achieve high compression rates while minimizing accuracy loss, providing a more efficient alternative to conventional uniform quantization methods.

Experience

University of California, Merced **Jan 2022 – present**
Research/Teaching Assistant *Merced, CA, United States*

- Research area: neural network compression (adaptive mix-bit quantization), learning oblique trees, tree-based methods for dimensionality reduction, clustering and regression problems.
- Assisting in teaching the following courses: Advanced Programming (Fall 2024), Introduction to OOP (Spring 2023), Introduction to programming (Fall 2022), Discrete Mathematics (Spring 2022)

Sberbank **Apr 2021 – Oct 2021**
NLP ML Engineer *Moscow, Russia*

- Participated in the development and fine-tuning of a NLP framework for Named Entity Recognition (NER) and classification, which supported different Large Language Models (LLMs) (BERT, RoBERTa, ALBERT, BiLM, and Longformer), enabling flexible model training.
- Trained the LayoutLM model from scratch for Russian-language document processing.
- Developed a custom parsing algorithm to extract entity positions from LabelStudio's NER task outputs, ensuring seamless compatibility with the framework.

Sberbank **Jun 2020 – Apr 2021**
Data Scientist (End-to-end model development) *Moscow, Russia*

- Developed an employee development model for the Sales Department, resulting in a 15% reduction in employee churn.
- Designed and developed a model to detect product imposition, helping the bank identify unethical practices.
- Created a product selling propensity model, improving customer targeting and increasing revenue by 10%.

Professional Activities

Reviewer: Journal of Machine Learning Research (JMLR 2023), Neural Information Processing Systems (NeurIPS 2024), ACM SIGKDD (KDD 2024), International Conference on Learning Representations (ICLR 2024)

Technical Skills

Languages: Python (PyTorch, scikit-learn, numpy, scipy, pandas, transformers, LightGBM, XGBoost), C/C++,
Technologies: Linux, Git, Matlab