1. CatBoost: unbiased boosting with categorical features

*Liudmila Prokhorenkova, Gleb Gusev, Aleksandr Vorobev, Anna Veronika Dorogush, Andrey Gulin. NeurIPS, 2018*

NeurIPS 2018 paper with explanation of Ordered boosting principles and ordered categorical features statistics.

1. CatBoost: gradient boosting with categorical features support

*Anna Veronika Dorogush, Vasily Ershov, Andrey Gulin. Workshop on ML Systems at NIPS 2017*

A paper explaining the CatBoost working principles: how it handles categorical features, how it fights overfitting, how GPU training and fast formula applier are implemented.

1. Minimal Variance Sampling in Stochastic Gradient Boosting

*Bulat Ibragimov, Gleb Gusev. arXiv:1910.13204*

A paper about Minimal Variance Sampling, which is the default sampling in CatBoost.

1. Finding Influential Training Samples for Gradient Boosted Decision Trees

*Boris Sharchilev, Yury Ustinovsky, Pavel Serdyukov, Maarten de Rijke. arXiv:1802.06640*

A paper explaining several ways of extending the framework for finding influential training samples for a particular case of tree ensemble-based models to non-parametric GBDT ensembles under the assumption that tree structures remain fixed and introducing a general scheme of obtaining further approximations to this method that balance the trade-off between performance and computational complexity.

1. A Unified Approach to Interpreting Model Predictions

*Scott Lundberg, Su-In Lee. arXiv:1705.07874*

A paper explaining a unified framework for interpreting predictions, SHAP (SHapley Additive exPlanations).

1. Consistent feature attribution for tree ensembles

*Scott M. Lundberg, Su-In Lee. arXiv:1706.06060*

A paper explaining fast exact solutions for SHAP (SHapley Additive exPlanation) values, a unique additive feature attribution method based on conditional expectations that is both consistent and locally accurate.

1. Winning The Transfer Learning Track of Yahoo!’s Learning To Rank Challenge with YetiRank

*Andrey Gulin, Igor Kuralenok, Dimitry Pavlov. PMLR 14:63-76*

The theory underlying the YetiRank and YetiRankPairwise modes in CatBoost.