Introduction to Machine Learning
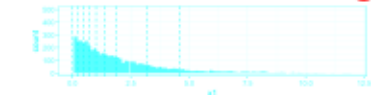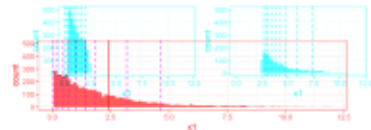
Gradient Boosting: XGBoost

Learning goals

- Overview over XGB
- Regularization in XGB
- Approximate split finding
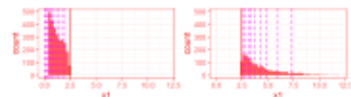
SUBSAMPLING

Data Subsampling: XGB uses stochastic GB.

Feature Subsampling: Similar to `mtry` in a random forest only a random subset of features is used for split finding.

The fraction of features for a split can be randomly sampled for each

❶ tree

❷ level of a tree

❸ split

Feature subsampling speeds up training even further and can create a more diverse ensemble that often performs better.

OVERVIEW OF IMPORTANT HYPERPARAMETERS

HP (as named in software)	Type	Typical Range	Trafo	Default	Description
eta	R	$[-4, 0]$	10^x	0.3	learning rate (also called ν) shrinks contribution of each boosting update
nrounds	I	$\{1, \ldots, 5000\}$	–	–	number of boosting iterations. Can also be optimized with early stopping.
gamma	R	$[-7, 6]$	2^x	0	minimum loss reduction required to make a further partition on a leaf node of the tree
max_depth	I	$\{1, \ldots, 20\}$	–	6	maximum depth of a tree
colsample_bytree	R	$[0.1, 1]$	–	1	subsample ratio of columns for each tree
colsample_bylevel	R	$[0.1, 1]$	–	1	subsample ratio of columns for each depth level
lambda	R	$[-10, 10]$	2^x	1	$L2$ regularization term on weights
alpha	R	$[-10, 10]$	2^x	0	$L1$ regularization term on weights
subsample	R	$[0.1, 1]$	–	1	subsample ratio of the training instances