

Performance estimation

Large dataset

- 2-way holdout method (train/test split)
- Confidence interval via normal approximation

Small dataset

- K-fold cross-validation (without independent test set)
- Leave-one-out cross-validation (without independent test set)
- Confidence interval via 0.632(+) bootstrap

Model selection (hyperparameter optimization) and performance estimation

Large dataset

- 3-way holdout method (train/validation/test split)

Small dataset

- K-fold cross-validation (with independent test set)
- Leave-one-out cross-validation (with independent test set)

Algorithm comparison

Large dataset

- McNemar test
- (Cochran's Q +) McNemar test

Small dataset

- K-fold cross-validated paired t test
- 5x2cv paired t test
- Nested cross-validation