

mlr3proba

Package website: [release](#) | [dev](#)

Probabilistic Supervised Learning for **mlr3**.

What is mlr3proba ?

mlr3proba is a probabilistic supervised learning (PSL) toolkit for machine learning in R utilising the **mlr3** package. Probabilistic supervised learning is a field of supervised machine learning in which probability distributions are predicted. Regression and classification tasks can be shown to be sub-fields of PSL, for example reducing a predicted probability distribution from probabilistic regression by taking its mean gives a regression prediction. PSL is therefore a powerful tool to provide more information about predictions than the more classical regression or classification. Probably the most known variant of PSL is survival analysis, where the task of interest is to predict an individual's survival curve. Other forms of PSL include density estimation and probabilistic regression. To-date, PSL toolkits in R have been limited to Bayesian simulation packages, but **mlr3proba** hopes to change this by allowing domain-agnostic (Bayesian or Frequentist) fit/predict and evaluation workflows.

Installation

Install the last release from CRAN:

```
install.packages("mlr3proba")
```

Install the development version from GitHub:

```
remotes::install_github("mlr-org/mlr3proba")
```

Survival Analysis

Survival Learners

ID	Learner	Package
surv.blackboost	Gradient Boosting with Regression Trees	mboost
surv.coxph	Cox Proportional Hazards	survival
surv.cvglmnet	Cross-Validated GLM with Elastic Net Regularization	glmnet
surv.flexible	Flexible Parametric Spline Models	flexsurv
surv.gamboost	Gradient Boosting for Additive Models	mboost
surv.gbm	Generalized Boosting Regression Modeling	gbm
surv.glmboost	Gradient Boosting with Component-wise Linear Models	mboost
surv.glmnet	GLM with Elastic Net Regularization	glmnet
surv.kaplan	Kaplan-Meier Estimator	survival
surv.mboost	Gradient Boosting for Generalized Additive Models	mboost
surv.nelson	Nelson-Aalen Estimator	survival
surv.parametric	Fully Parametric Survival Models	survival
surv.penalized	L1 and L2 Penalized Estimation in GLMs	penalized
surv.randomForestSRC	RandomForestSRC Survival Forest	randomForestSRC

ID	Learner	Package
surv.ranger	Ranger Survival Forest	ranger
surv.rpart	Rpart Survival Forest	rpart
surv.svm	Regression, Ranking and Hybrid Support Vector Machines	survivalsvm

Survival Measures

ID	Learner	Package
surv.beggC	Begg's C-Index	survAUC
surv.chamblessAUC	Chambless and Diao's AUC	survAUC
surv.gonenC	Gonen and Heller's C-Index	survAUC
surv.graf	Integrated Graf Score	mlr3proba
surv.grafSE	Standard Error of Integrated Graf Score	mlr3proba
surv.harrellC	Harrell's C-Index	mlr3proba
surv.hungAUC	Hung and Chiang's AUC	survAUC
surv.intlogloss	Integrated Log Loss	mlr3proba
surv.intloglossSE	Standard Error of Integrated Log Loss	mlr3proba
surv.logloss	Log Loss	mlr3proba
surv.loglossSE	Standard Error of Log Loss	mlr3proba
surv.nagelkR2	Nagelkerke's R2	survAUC
surv.oquigleyR2	O'Quigley, Xu, and Stare's R2	survAUC
surv.songAUC	Song and Zhou's AUC	survAUC
surv.songTNR	Song and Zhou's TNR	survAUC
surv.songTPR	Song and Zhou's TPR	survAUC
surv.unoAUC	Uno's AUC	survAUC
surv.unoC	Uno's C-Index	survAUC
surv.unoTNR	Uno's TNR	survAUC
surv.unoTPR	Uno's TPR	survAUC
surv.xuR2	Xu and O'Quigley's R2	survAUC

Feature Overview and Lifecycle

The vision of **mlr3proba** is to be the first complete probabilistic machine learning package in R. This encompasses survival analysis, probabilistic regression, and unsupervised density estimation. The first release of **mlr3proba** is focused entirely on survival analysis and introduces **TaskSurv**. Later releases will include **TaskDensity** and will extend **TaskRegr** to have probabilistic predict types. The lifecycle of the survival task and features are considered **maturing** and any major changes are unlikely. The density and probabilistic regression tasks are currently in the early stages of development. The current main features of **mlr3proba** are:

- The added **TaskSurv**, **LearnerSurv**, **PredictionSurv** for survival analysis
- 17 survival learners, and 21 survival measures, including efficient implementations of censoring-adjusted probabilistic measures, such as the Integrated Graf (or Brier) Score.
- PipeOps integrated with **mlr3pipelines** for composition of probability distributions from linear predictors

Future Plans

- Add **TaskDensity**, **PredictionDensity**, **LearnerDensity**, and associated learners/measures

- Add **prob** predict type to **TaskRegr**, and associated learners/measures
- Allow **MeasureSurv** to return measures at multiple time-points simultaneously
- Improve estimation of integrated scores, and re-implement **survAUC** scores in **mlr3proba**
- Continue to add survival measures and learners

Bugs, Questions, Feedback

mlr3proba is a free and open source software project that encourages participation and feedback. If you have any issues, questions, suggestions or feedback, please do not hesitate to open an “issue” about it on the GitHub page!

In case of problems / bugs, it is often helpful if you provide a “minimum working example” that showcases the behaviour (but don’t worry about this if the bug is obvious).

Please understand that the resources of the project are limited: response may sometimes be delayed by a few days, and some feature suggestions may be rejected if they are deemed too tangential to the vision behind the project.

Similar Projects

A predecessor to this package is **mlr**, using the survival task. Several packages exist for pure Bayesian probabilistic modelling, including **jags** and **stan**. For implementation of a few survival models and measures, the largest package is **survival**. There does not appear to be a package that implements many different variants of density estimation, but see **this list** for the biggest density estimation packages in R.