

Super Learning (SL) and sl3

Rachael Phillips

rachaelvphillips@berkeley.edu

JSM 2023

Overview of Super Learner

Super Learner (SL)

LIBRARY

Linear model

BART

Random Forest

Neural
Network

Lasso

HAL

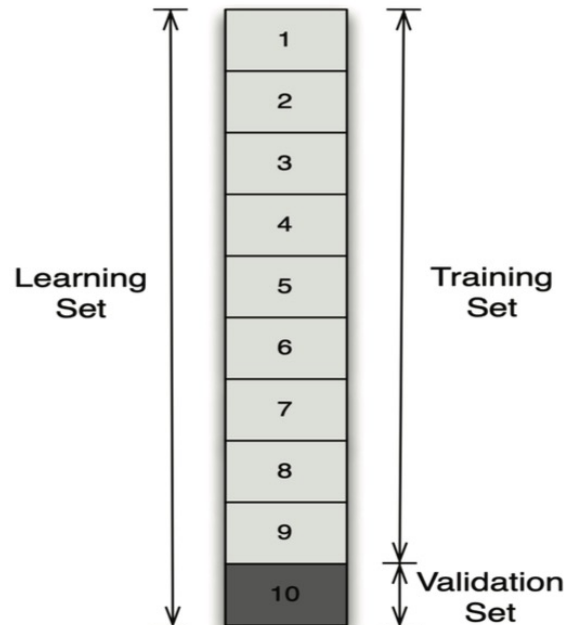
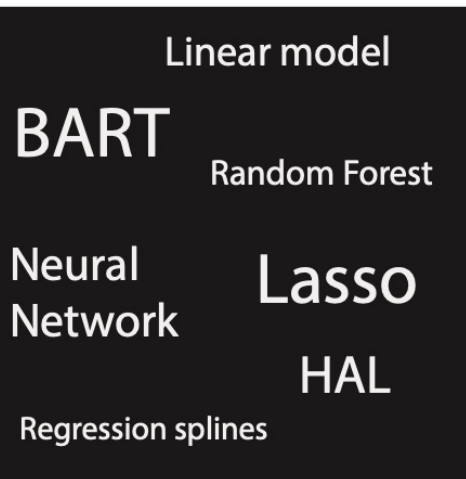
Regression splines

Super Learner (SL)

LIBRARY

COMPETITION

Cross-validated
performance of
learners + ensembles



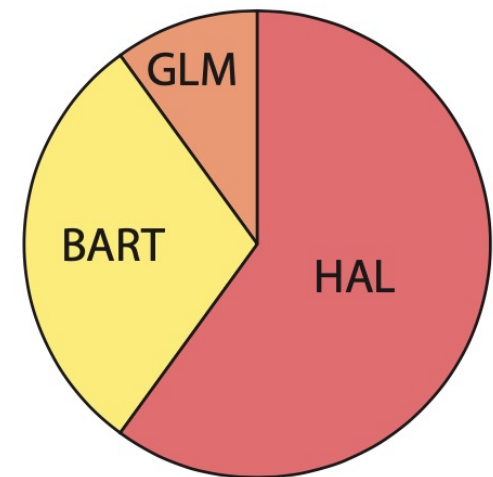
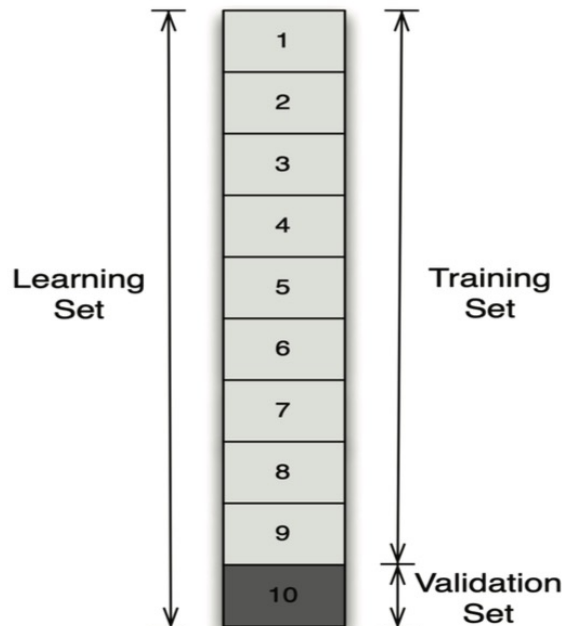
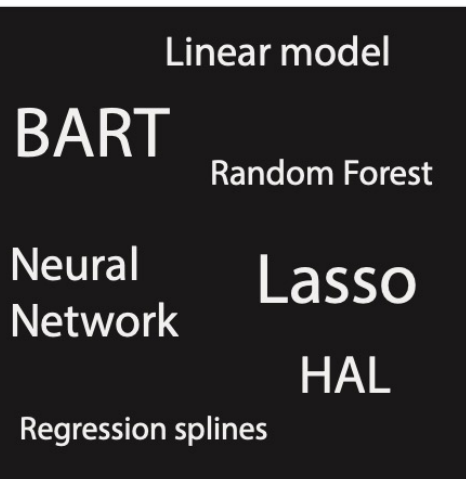
Super Learner (SL)

LIBRARY

COMPETITION

WINNER

Cross-validated
performance of
learners + ensembles



Super Learner (SL)

- Uses a library of algorithms for estimating a prediction function
 - Analyst specifies Alg_1, \dots, Alg_K
 - Create an optimal combination
 - Optimal with respect to V-fold cross-validated (CV) risk
 - Example risk functions: Negative log likelihood, mean squared error, 1-AUC
- SL predicted values, \hat{Y}_{SL} , are a combination of $\hat{Y}_{Alg_1}, \dots, \hat{Y}_{Alg_K}$
 - Discrete SL: “winner-take-all”, predictions from algorithm with best CV risk
 - Ensemble SL: predictions from multiple algorithms are combined
 - weighted combination
 - some other, possibly complex function of the algorithms’ predictions

Why super learner (SL)?

- No need to select the one “right” strategy
- Can consider diverse set
- Grounded in statistical optimality theory
- Pre-specified also flexible
- Mitigate statistical model misspecification
- Good initial estimation for estimators common for causal inference (e.g., AIPW, TMLE)

What does user specify for SL?

- a) Measure of performance
- b) Cross-validation scheme
- c) Diverse library of candidate learners

Super learner

1. Specify

- a) Measure of performance
- b) Cross-validation scheme
- c) Diverse library of candidate learners

What are you learning from the data? What do you want to optimize for?

Performance measure should be valid (i.e., optimized by underlying target), bounded, corresponds to desired goal

Super learner

1. Specify

- a) Measure of performance
- b) Cross-validation scheme
- c) Diverse library of candidate learners

e.g. V-fold cross-validation

1	1	1	1	1	1	1	1	1	1
2	2	2	2	2	2	2	2	2	2
3	3	3	3	3	3	3	3	3	3
4	4	4	4	4	4	4	4	4	4
5	5	5	5	5	5	5	5	5	5
6	6	6	6	6	6	6	6	6	6
7	7	7	7	7	7	7	7	7	7
8	8	8	8	8	8	8	8	8	8
9	9	9	9	9	9	9	9	9	9
10	10	10	10	10	10	10	10	10	10
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5	Fold 6	Fold 7	Fold 8	Fold 9	Fold 10

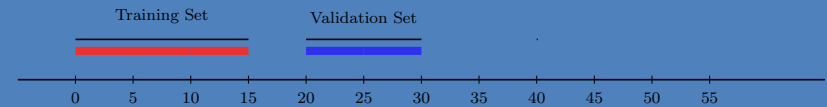
Super learner

1. Specify

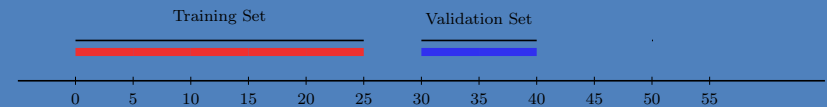
- Measure of performance
- Cross-validation scheme
- Diverse library of candidate learners

e.g. rolling origin cross-validation
(cross-validation for time series data)

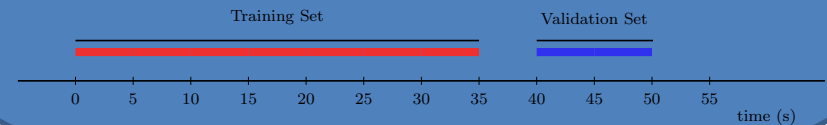
$v = 1$



$v = 2$



$v = 3$

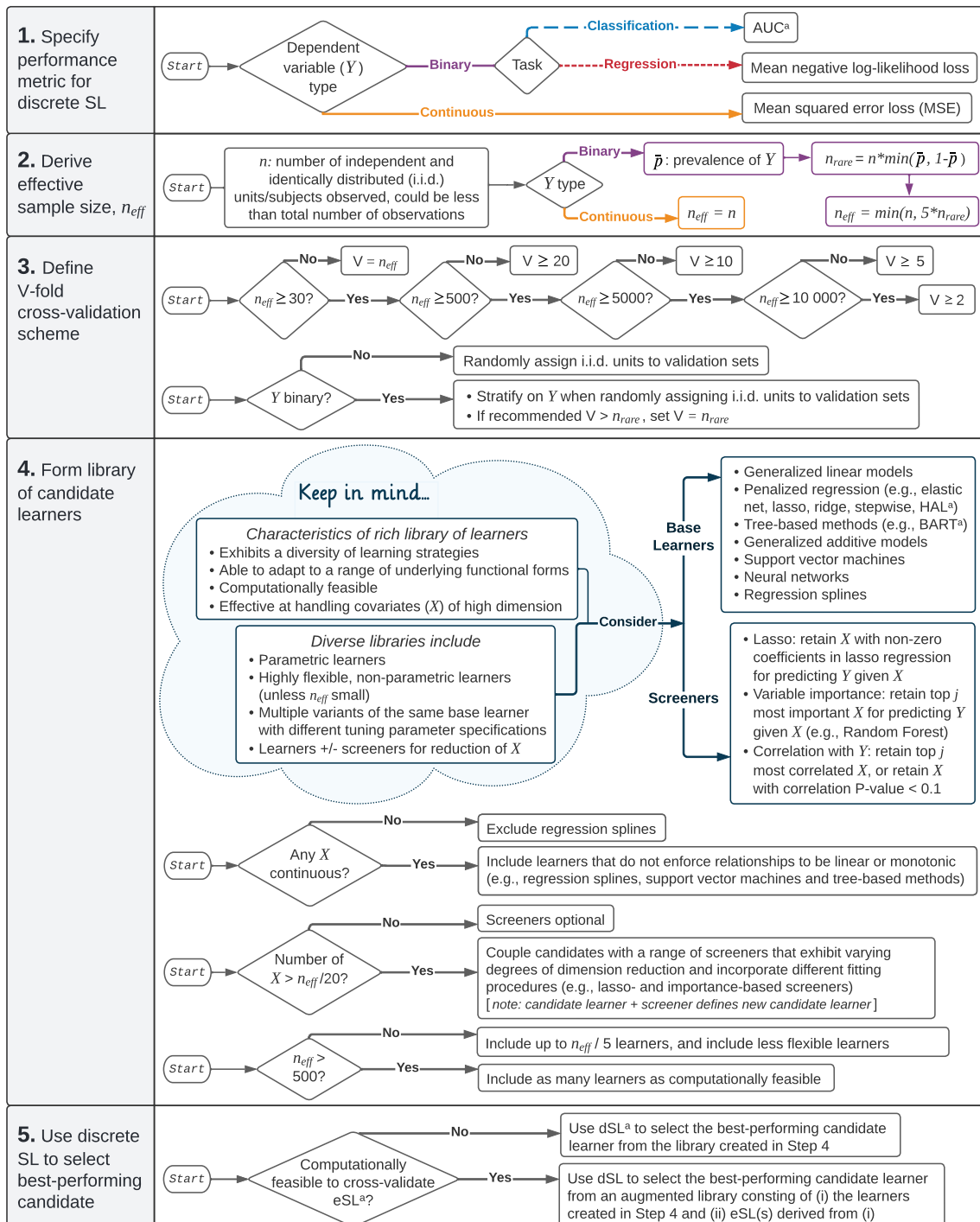


Overview of the algorithm

1. Make metalevel dataset with cross-validated candidate predictions and validation set outcomes
2. Fit meta-learner to the metalevel dataset
3. Full-fit candidates
4. Define the SL

Practical considerations for specifying SL

Rachael V Phillips, Mark J van der Laan, Hana Lee, Susan Gruber, Practical considerations for specifying a super learner, *International Journal of Epidemiology*, 2023.
<https://doi.org/10.1093/ije/dyad023>



sl3

SL software package in tlverse

Introductory overview of s13

- Task
- Learners
- Functions

Introductory overview of s13

- Task
- Learners
- Functions

Tasks in sl3

What is the prediction task?

data, covariates, outcome,
weights, id, outcome_type, offset,
drop_missing_outcome, folds

https://tlverse.org/sl3/reference/sl3_Task.html

Introductory overview of s13

- Task
- **Learners**
- Functions

Exercise: Training learner with `s13` wrapper versus package's function

- Prediction tasks in `s13`
- Comparison of R6 methods and S3 methods
 - `s13` R package function, `Lrnr_earth`
 - `earth` R package function, `earth`
- Looking up learners, the `s13` wrapper, and arguments
- Modifying learner parameters
- **Data example:** Collaborative Perinatal Project (CPP) was a multisite prospective cohort study designed to identify the effects of complications during either pregnancy or the perinatal period on birth and child outcomes.

Learners in sl3

How to estimate prediction function?

Learners in sl3

How to estimate prediction function?

Other Learners: `Lrnr_HarmonicReg`, `Lrnr_arima`, `Lrnr_bartMachine`, `Lrnr_base`,
`Lrnr_bayesglm`, `Lrnr_bilstm`, `Lrnr_caret`, `Lrnr_cv_selector`, `Lrnr_cv`, `Lrnr_dbarts`,
`Lrnr_define_interactions`, `Lrnr_density_discretize`, `Lrnr_density_hse`,
`Lrnr_density_semiparametric`, `Lrnr_earth`, `Lrnr_expSmooth`, `Lrnr_gam`, `Lrnr_ga`,
`Lrnr_gbm`, `Lrnr_glm_fast`, `Lrnr_glmnet`, `Lrnr_glm`, `Lrnr_grf`, `Lrnr_gru_keras`, `Lrnr_gts`,
`Lrnr_h2o_grid`, `Lrnr_hal9001`, `Lrnr_haldensify`, `Lrnr_hts`, `Lrnr_independent_binomial`,
`Lrnr_lightgbm`, `Lrnr_lstm_keras`, `Lrnr_mean`, `Lrnr_multiple_ts`, `Lrnr_multivariate`,
`Lrnr_nnet`, `Lrnr_nnls`, `Lrnr_optim`, `Lrnr_pca`, `Lrnr_pkg_SuperLearner`, `Lrnr_polspline`,
`Lrnr_pooled_hazards`, `Lrnr_randomForest`, `Lrnr_ranger`, `Lrnr_revere_task`, `Lrnr_rpart`,
`Lrnr_rugarch`, `Lrnr_screener_augment`, `Lrnr_screener_coefs`,
`Lrnr_screener_correlation`, `Lrnr_screener_importance`, `Lrnr_sl`, `Lrnr_solnp_density`,
`Lrnr_solnp`, `Lrnr_stratified`, `Lrnr_subset_covariates`, `Lrnr_svm`, `Lrnr_tsDyn`,
`Lrnr_ts_weights`, `Lrnr_xgboost`, `Pipeline`, `Stack`, `define_h2o_X()`,

Learners in sl3

How to estimate prediction function?

Other Learners: `Lrnr_HarmonicReg`, `Lrnr_arima`, `Lrnr_bartMachine`, `Lrnr_base`,
`Lrnr_bayesglm`, `Lrnr_bilstm`, `Lrnr_caret`, `Lrnr_cv_selector`, `Lrnr_cv`, `Lrnr_dbarts`,
`Lrnr_define_interactions`, `Lrnr_density_discretize`, `Lrnr_density_hse`,
`Lrnr_density_semiparametric`, `Lrnr_earth`, `Lrnr_expSmooth`, `Lrnr_gam`, `Lrnr_ga`,
`Lrnr_gbm`, `Lrnr_glm_fast`, `Lrnr_glmnet`, `Lrnr_glm`, `Lrnr_grf`, `Lrnr_gru_keras`, `Lrnr_gts`,
`Lrnr_h2o_grid`, `Lrnr_hal9001`, `Lrnr_haldensify`, `Lrnr_hts`, `Lrnr_independent_binomial`,
`Lrnr_lightgbm`, `Lrnr_lstm_keras`, `Lrnr_mean`, `Lrnr_multiple_ts`, `Lrnr_multivariate`,
`Lrnr_nnet`, `Lrnr_nnls`, `Lrnr_optim`, `Lrnr_pca`, `Lrnr_pkg_SuperLearner`, `Lrnr_polspline`,
`Lrnr_pooled_hazards`, `Lrnr_randomForest`, `Lrnr_ranger`, `Lrnr_revere_task`, `Lrnr_rpart`,
`Lrnr_rugarch`, `Lrnr_screener_augment`, `Lrnr_screener_coefs`,
`Lrnr_screener_correlation`, `Lrnr_screener_importance`, `Lrnr_sl`, `Lrnr_solnp_density`,
`Lrnr_solnp`, `Lrnr_stratified`, `Lrnr_subset_covariates`, `Lrnr_svm`, `Lrnr_tsDyn`,
`Lrnr_ts_weights`, `Lrnr_xgboost`, `Pipeline`, `Stack`, `define_h2o_X()`,

Learners in sl3

How to estimate prediction function?

Other Learners: Lrnr_HarmonicReg, Lrnr_arima, Lrnr_bartMachine, Lrnr_base, Lrnr_bayesglm, Lrnr_bilstm, Lrnr_caret, Lrnr_cv_selector, Lrnr_cv, Lrnr_dbarts, Lrnr_define_interactions, Lrnr_density_discretize, Lrnr_density_hse, Lrnr_density_semiparametric, Lrnr_earth, Lrnr_expSmooth, Lrnr_gam, Lrnr_ga, Lrnr_gbm, Lrnr_glm_fast, Lrnr_glmnet, Lrnr_glm, Lrnr_grf, Lrnr_gru_keras, Lrnr_gts, Lrnr_h2o_grid, Lrnr_hal9001, Lrnr_haldensify, Lrnr_hts, Lrnr_independent_binomial, Lrnr_lightgbm, Lrnr_lstm_keras, Lrnr_mean, Lrnr_multiple_ts, Lrnr_multivariate, Lrnr_nnet, Lrnr_nnls, Lrnr_optim, Lrnr_pca, Lrnr_pkg_SuperLearner, Lrnr_polspline, Lrnr_pooled_hazards, Lrnr_randomForest, Lrnr_ranger, Lrnr_revere_task, Lrnr_rpart, Lrnr_rugarch, Lrnr_screener_augment, Lrnr_screener_coefs, Lrnr_screener_correlation, Lrnr_screener_importance, Lrnr_sl, Lrnr_solnp_density, Lrnr_solnp, Lrnr_stratified, Lrnr_subset_covariates, Lrnr_svm, Lrnr_tsDyn, Lrnr_ts_weights, Lrnr_xgboost, Pipeline, Stack, define_h2o_X(),

Learners in sl3

How to estimate prediction function?

Other Learners: Lrnr_HarmonicReg, Lrnr_arima, Lrnr_bartMachine, Lrnr_base, Lrnr_bayesglm, Lrnr_bilstm, Lrnr_caret, Lrnr_cv_selector, Lrnr_cv, Lrnr_dbarts, Lrnr_define_interactions, Lrnr_density_discretize, Lrnr_density_hse, Lrnr_density_semiparametric, Lrnr_earth, Lrnr_expSmooth, Lrnr_gam, Lrnr_ga, Lrnr_gbm, Lrnr_glm_fast, Lrnr_glmnet, Lrnr_glm, Lrnr_grf, Lrnr_gru_keras, Lrnr_gts, Lrnr_h2o_grid, Lrnr_hal9001, Lrnr_haldensify, Lrnr_hts, Lrnr_independent_binomial, Lrnr_lightgbm, Lrnr_lstm_keras, Lrnr_mean, Lrnr_multiple_ts, Lrnr_multivariate, Lrnr_nnet, Lrnr_nnls, Lrnr_optim, Lrnr_pca, Lrnr_pkg_SuperLearner, Lrnr_polspline, Lrnr_pooled_hazards, Lrnr_randomForest, Lrnr_ranger, Lrnr_revere_task, Lrnr_rpart, Lrnr_rugarch, Lrnr_screener_augment, Lrnr_screener_coefs, Lrnr_screener_correlation, Lrnr_screener_importance, Lrnr_sl, Lrnr_solnp_density, Lrnr_solnp, Lrnr_stratified, Lrnr_subset_covariates, Lrnr_svm, Lrnr_tsDyn, Lrnr_ts_weights, Lrnr_xgboost, Pipeline, **Stack**, define_h2o_X(),

Learners in sl3

How to estimate prediction function?

Other Learners: Lrnr_HarmonicReg, Lrnr_arma, Lrnr_bartMachine, Lrnr_base, Lrnr_bayesglm, Lrnr_bilstm, Lrnr_caret, Lrnr_cv_selector, Lrnr_cv, Lrnr_dbarts, Lrnr_define_interactions, Lrnr_density_discretize, Lrnr_density_hse, Lrnr_density_semiparametric, Lrnr_earth, Lrnr_expSmooth, Lrnr_gam, Lrnr_ga, Lrnr_gbm, Lrnr_glm_fast, Lrnr_glmnet, Lrnr_glm, Lrnr_grf, Lrnr_gru_keras, Lrnr_gts, Lrnr_h2o_grid, Lrnr_hal9001, Lrnr_haldensify, Lrnr_hts, Lrnr_independent_binomial, Lrnr_lightgbm, Lrnr_lstm_keras, Lrnr_mean, Lrnr_multiple_ts, Lrnr_multivariate, Lrnr_nnet, Lrnr_nnls, Lrnr_optim, Lrnr_pca, Lrnr_pkg_SuperLearner, Lrnr_polspline, Lrnr_pooled_hazards, Lrnr_randomForest, Lrnr_ranger, Lrnr_revere_task, Lrnr_rpart, Lrnr_rugarch, Lrnr_screener_augment, Lrnr_screener_coefs, Lrnr_screener_correlation, Lrnr_screener_importance, Lrnr_sl, Lrnr_solnp_density, Lrnr_solnp, Lrnr_stratified, Lrnr_subset_covariates, Lrnr_svm, Lrnr_tsDyn, Lrnr_ts_weights, Lrnr_xgboost, Pipeline, Stack, define_h2o_X(),

A diagram consisting of five red ovals and one orange rectangle. The ovals are positioned around the text: one around 'Lrnr_caret', one around 'Lrnr_glm', one around 'Lrnr_gam', one around 'Lrnr_ranger', and one around 'Lrnr_ts_weights'. The orange rectangle is positioned around the word 'Stack'. Red lines connect each of the five ovals to the orange rectangle, indicating that these learners are components or related to the 'Stack' learner.

Learners in sl3

How to estimate prediction function?

Other Learners: `Lrnr_HarmonicReg`, `Lrnr_arima`, `Lrnr_bartMachine`, `Lrnr_base`,
`Lrnr_bayesglm`, `Lrnr_bilstm`, `Lrnr_caret`, `Lrnr_cv_selector`, `Lrnr_cv`, `Lrnr_dbarts`,
`Lrnr_define_interactions`, `Lrnr_density_discretize`, `Lrnr_density_hse`,
`Lrnr_density_semiparametric`, `Lrnr_earth`, `Lrnr_expSmooth`, `Lrnr_gam`, `Lrnr_ga`,
`Lrnr_gbm`, `Lrnr_glm_fast`, `Lrnr_glmnet`, `Lrnr_glm`, `Lrnr_grf`, `Lrnr_gru_keras`, `Lrnr_gts`,
`Lrnr_h2o_grid`, `Lrnr_hal9001`, `Lrnr_haldensify`, `Lrnr_hts`, `Lrnr_independent_binomial`,
`Lrnr_lightgbm`, `Lrnr_lstm_keras`, `Lrnr_mean`, `Lrnr_multiple_ts`, `Lrnr_multivariate`,
`Lrnr_nnet`, `Lrnr_nnls`, `Lrnr_optim`, `Lrnr_pca`, `Lrnr_pkg_SuperLearner`, `Lrnr_polspline`,
`Lrnr_pooled_hazards`, `Lrnr_randomForest`, `Lrnr_ranger`, `Lrnr_revere_task`, `Lrnr_rpart`,
`Lrnr_rugarch`, `Lrnr_screener_augment`, `Lrnr_screener_coefs`,
`Lrnr_screener_correlation`, `Lrnr_screener_importance`, `Lrnr_sl`, `Lrnr_solnp_density`,
`Lrnr_solnp`, `Lrnr_stratified`, `Lrnr_subset_covariates`, `Lrnr_svm`, `Lrnr_tsDyn`,
`Lrnr_ts_weights`, `Lrnr_xgboost`, `Pipeline`, `Stack`, `define_h2o_X()`,

Learners in sl3

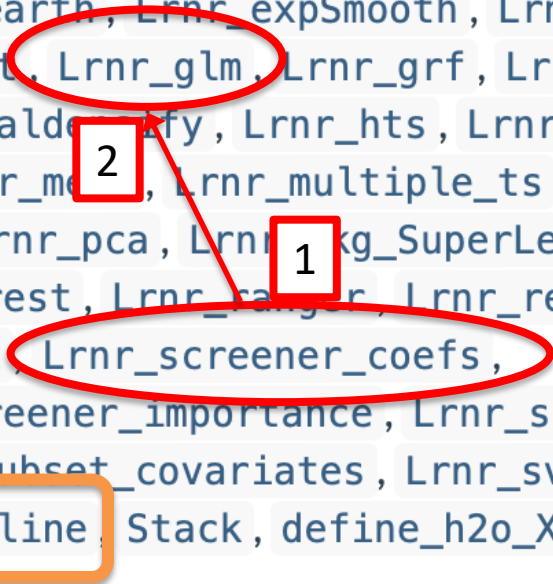
How to estimate prediction function?

Other Learners: `Lrnr_HarmonicReg`, `Lrnr_arima`, `Lrnr_bartMachine`, `Lrnr_base`,
`Lrnr_bayesglm`, `Lrnr_bilstm`, `Lrnr_caret`, `Lrnr_cv_selector`, `Lrnr_cv`, `Lrnr_dbarts`,
`Lrnr_define_interactions`, `Lrnr_density_discretize`, `Lrnr_density_hse`,
`Lrnr_density_semiparametric`, `Lrnr_earth`, `Lrnr_expSmooth`, `Lrnr_gam`, `Lrnr_ga`,
`Lrnr_gbm`, `Lrnr_glm_fast`, `Lrnr_glmnet`, `Lrnr_glm`, `Lrnr_grf`, `Lrnr_gru_keras`, `Lrnr_gts`,
`Lrnr_h2o_grid`, `Lrnr_hal9001`, `Lrnr_haldensify`, `Lrnr_hts`, `Lrnr_independent_binomial`,
`Lrnr_lightgbm`, `Lrnr_lstm_keras`, `Lrnr_mean`, `Lrnr_multiple_ts`, `Lrnr_multivariate`,
`Lrnr_nnet`, `Lrnr_nnls`, `Lrnr_optim`, `Lrnr_pca`, `Lrnr_pkg_SuperLearner`, `Lrnr_polspline`,
`Lrnr_pooled_hazards`, `Lrnr_randomForest`, `Lrnr_ranger`, `Lrnr_revere_task`, `Lrnr_rpart`,
`Lrnr_rugarch`, `Lrnr_screener_augment`, `Lrnr_screener_coefs`,
`Lrnr_screener_correlation`, `Lrnr_screener_importance`, `Lrnr_sl`, `Lrnr_solnp_density`,
`Lrnr_solnp`, `Lrnr_stratified`, `Lrnr_subset_covariates`, `Lrnr_svm`, `Lrnr_tsDyn`,
`Lrnr_ts_weights`, `Lrnr_xgboost`, `Pipeline`, `Stack`, `define_h2o_X()`,

Learners in sl3

How to estimate prediction function?

Other Learners: Lrnr_HarmonicReg, Lrnr_arima, Lrnr_bartMachine, Lrnr_base, Lrnr_bayesglm, Lrnr_bilstm, Lrnr_caret, Lrnr_cv_selector, Lrnr_cv, Lrnr_dbarts, Lrnr_define_interactions, Lrnr_density_discretize, Lrnr_density_hse, Lrnr_density_semiparametric, Lrnr_earth, Lrnr_expSmooth, Lrnr_gam, Lrnr_ga, Lrnr_gbm, Lrnr_glm_fast, Lrnr_glmnet, Lrnr_glm, Lrnr_grf, Lrnr_gru_keras, Lrnr_gts, Lrnr_h2o_grid, Lrnr_hal9001, Lrnr_haldane, Lrnr_hf, Lrnr_hts, Lrnr_independent_binomial, Lrnr_lightgbm, Lrnr_lstm_keras, Lrnr_m, Lrnr_multiple_ts, Lrnr_multivariate, Lrnr_nnet, Lrnr_nnls, Lrnr_optim, Lrnr_pca, Lrnr_pkg_SuperLearner, Lrnr_polspline, Lrnr_pooled_hazards, Lrnr_randomForest, Lrnr_sanger, Lrnr_revere_task, Lrnr_rpart, Lrnr_rugarch, Lrnr_screener_augment, Lrnr_screener_coefs, Lrnr_screener_correlation, Lrnr_screener_importance, Lrnr_sl, Lrnr_solnp_density, Lrnr_solnp, Lrnr_stratified, Lrnr_subset_covariates, Lrnr_svm, Lrnr_tsDyn, Lrnr_ts_weights, Lrnr_xgboost, Pipeline, Stack, define_h2o_X(),



Introductory overview of s13

- Task
- Learners
- **Other functions**

Other `sl3` Functions

- Performance measures:
 - loss functions (e.g., squared error, negative log-likelihood, multinomial log-likelihood)
 - metrics based on ROCR software package, like AUC, AUCPR, accuracy, sensitivity, with `custom_ROCR_risk()`
- Variable importance with `importance`
- Table with each candidate learner's cross-validated predictive performance with `cv_risk`
- Cross-validated SL (to see CV risk of SL itself) with `cv_sl`

Live coding exercise with sl3 using WASH Benefits data

WASH Benefits Bangladesh

- Study aiming to understand the **impact of water quality, sanitation, hand washing, and nutritional (WASH) interventions on child development in rural Bangladesh** (WASH Benefits Bangladesh): a cluster randomized controlled trial (Tofail et al. [2018](#))
- Enrolled pregnant women in first or second trimester from the rural villages of Bangladesh

WASH Benefits Bangladesh: Example Dataset for sl3 Prediction

- We concentrate on child growth (size for age) as the outcome of interest.
- We will use SL (sl3 software) to estimate the **prediction function** for predicting a child's weight-for-height z-score (whz) given covariate information.
- Next section on TMLE (tmle3 software) focuses on estimation of the (causal) impact of particular intervention on an outcome (e.g., the impact of the WASH intervention on whz)

Exercise: Training a super learner with sl_3

<https://tlverse.org/tlverse-handbook/sl3.html>

<http://tlverse.org/jsm2023-workshop/sl3.html>