LONDON SCHOOL of HYGIENE & TROPICAL MEDICINE


# Return of Super Learner

## David Whitney*

## Machine Learning 2022

*based on Prof Karla Diaz-Ordaz's 2021 materials

# Learning Outcomes

After completing today's session, you will

- understand the basic idea of Super Learner (SL)
- understand the 3 important choices to be made for using SL
- be able to apply SL in a practical example
- know how to use cross-validation to measure the performance of a SL

# Why Super Learner?

# Optimality: the oracle property

- We define the ‘oracle’ selector as estimator that minimises the risk under the true data-generating distribution.
- SL has same expected risk difference as the oracle selector (up to a second order term) asymptotically
- i.e. we do not ‘lose’ anything by using the SL even if we know a situation is best suited to a particular estimator (just add it in the library!)

# Conditions

- $L$, number of candidate learners in the library can grow at most polynomial in sample size $n$
    - to avoid over-fitting: estimated risk is close to true expected loss
- SL rate of convergence is $\log(n)/n$ (for a sufficiently large sample size) if it includes a parametric model which includes the truth

# Empirical performance

# Performance

- SL is a (meta-)learner (e.g runs a meta NNLS regression) as such, its performance should be studied via CV
- Computer intensive:
  - we can parallelise it (extra material computer lab)
  - we can run a scalable version (not covered)

# Cross-validated Super Learner

- Estimate cross-validated risk of the SL itself to study the SL performance
- This requires an "external" CV layer (nested CV)
- Requires setting aside a separate holdout sample not used to fit the SL, and test the SL predictions on it
- A loss function to evaluate SL needs to be specified
- This external CV procedure incorporate $F$ folds
- In the lab today, we use 3-5 outer/external folds of CV, usually 10.

# Time for some more practice!

# Interpretability

- Super Learner is a "black box" algorithm
  - The contribution of each variable is unclear
- Obtain variable importance metrics: quantifying marginal association between each predictor and outcome after adjusting for the others
  - A large contribution means strongly associated (can be confounding)

# Advanced topics (not covered today)

- Select a loss function that is appropriate for the parameter to be estimated (change the method in R function)
- Variable importance for the SL and using this to select the learners
- Feature screening: Select a subset of available covariates and pass only those variables to the modelling algorithm.

# Summary

# What you have learnt

- Super Learner can help us choose the single best learner
- But usually, we want the best weighted combination
- ML models are not necessarily going to be always better than parametric models
    - ...but we don't usually know when there is a gain
- Key contribution of SL is that it is guaranteed to attain the performance of the best algorithm considered (oracle property)
- SL helps with transparency by pre-specifying the analysis strategy while being data-adaptive
    - attenuates model mis-specification
    - uses CV to avoid over-fitting
- Important choices : loss functions and meta-learner, library

# SuperLearner, the R package

- Allows different loss functions and meta-learner functions (via `method`)
- Nested CV built-in
- Many machine algorithms have been included (with default hyper-parameters!!)
- CV can be used to tune these
- Parallel implementations
- Several other R implementations available (e.g. sl3 in active development)
- Python implementation in `mlens` (I have not played with this, though!)

# Links to practical guides and alternative software

- https://cran.r-project.org/web/packages/SuperLearner/vignettes/Guide-to-SuperLearner.html
- https://tlverse.org/tlverse-handbook/sl3.html
- http://ml-ensemble.com (Python)

# Further reading and references


- van der Laan, M. J., and Rose, S. (2011) Targeted learning: causal inference for observational and experimental data. Springer Science & Business Media (Ch 3)

- van der Laan, MJ and Rose S. (2018) Targeted Learning in Data Science, Springer Series in Statistics (Ch 18)

- Polley, E. C. and van der Laan, M. J. (2010) Super learner in prediction. U.C. Berkeley Division of Biostatistics Working Paper Series Working Paper 266. URL http://biostats.bepress.com/ucbbiostat/paper266.

- Naimi A and Balzer L. Stacked generalization: an introduction to super learning. European Journal of Epidemiology (2018) 33:459–464

- Bühlmann, P. (Ed.), Drineas, P. (Ed.), Kane, M. (Ed.), van der Laan, M. (Ed.). (2016). Handbook of Big Data. New York: Chapman and Hall/CRC, (Chapter 19)