Consistency between ML and classical approaches for covariate identification

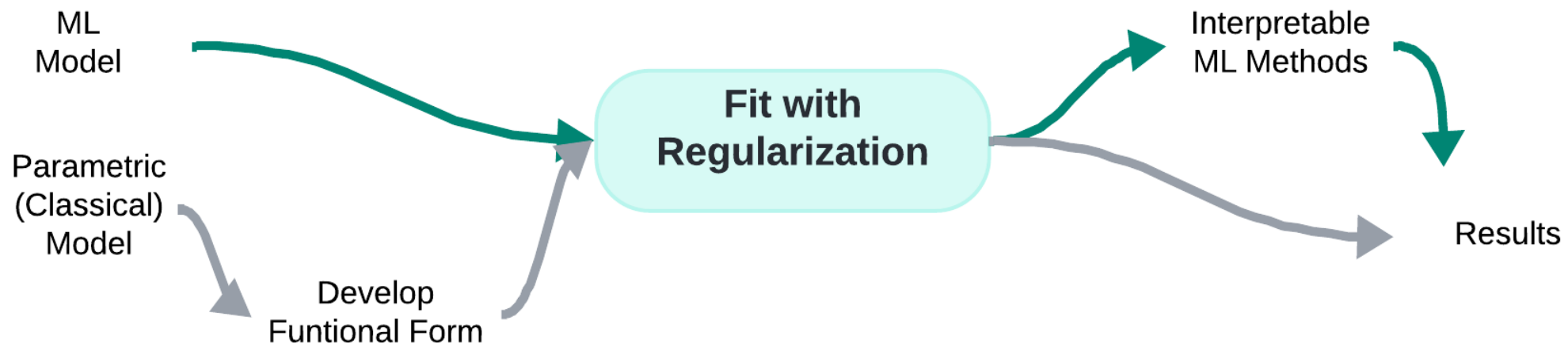
Matthew Wiens
March 26, 2024

METRUM
RESEARCH GROUP

ASCPT 2024 ANNUAL MEETING

BIG IDEA

- Machine Learning ideas (e.g. regularization) applied to other models is a powerful combination

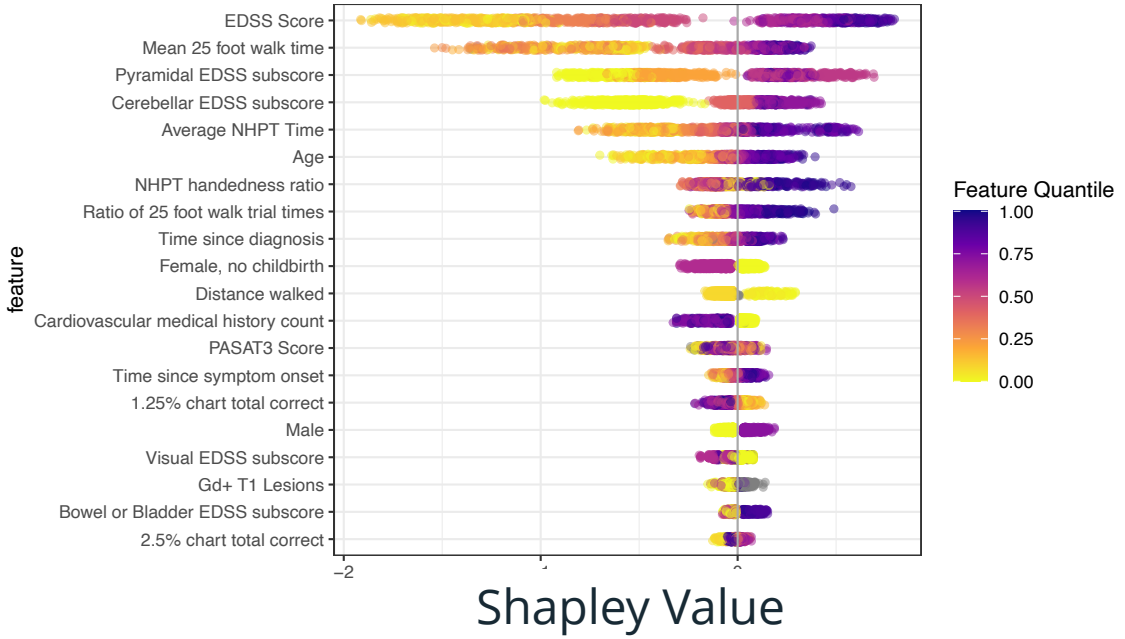


ML WORKFLOW: REGULARIZATION *AND* INTERPRETABLE AI

- Assemble data and choose model structure
- Tune (regularizing) hyperparameters for complex data
  - Large number of covariates
  - Complex relationships (nonlinear, interactions)
  - Non-standard data types (-omics)
- Interpret model fit
  - Shapley values
  - Variable importance
  - Predictions

SHAPLEY VALUES FOR COVARIATE INTERPRETATION

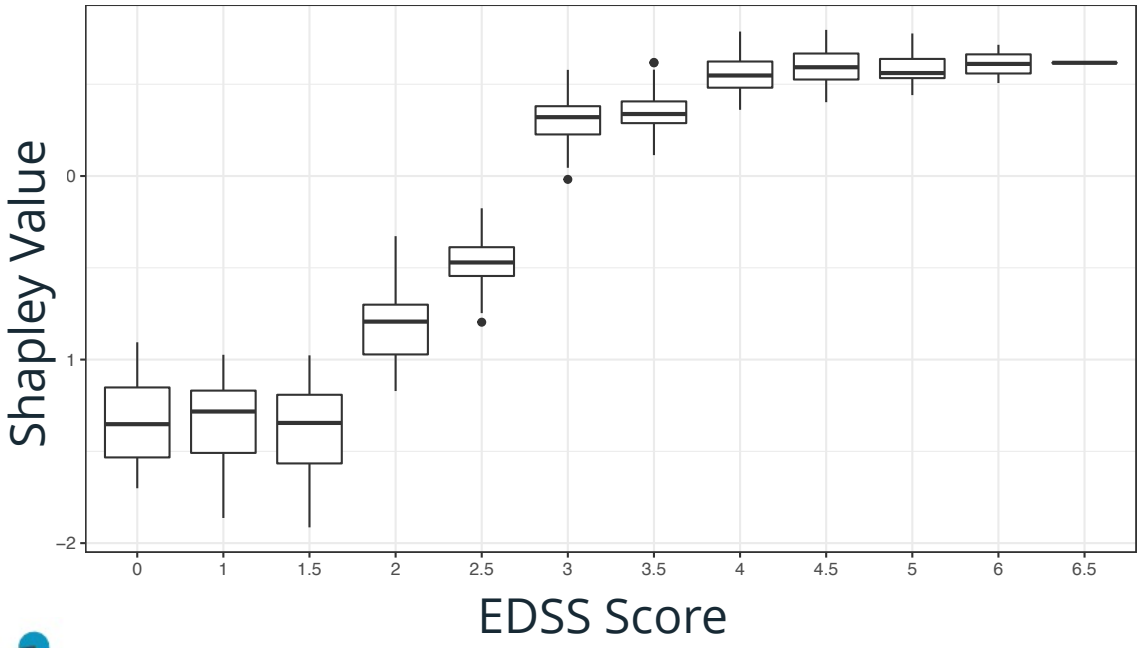
- Idea: Value of players (features) drafted onto sports team (model)
- Characterize:
  - magnitude
  - variability
  - shape
- Example: Predicting changes in subtype of Multiple Sclerosis

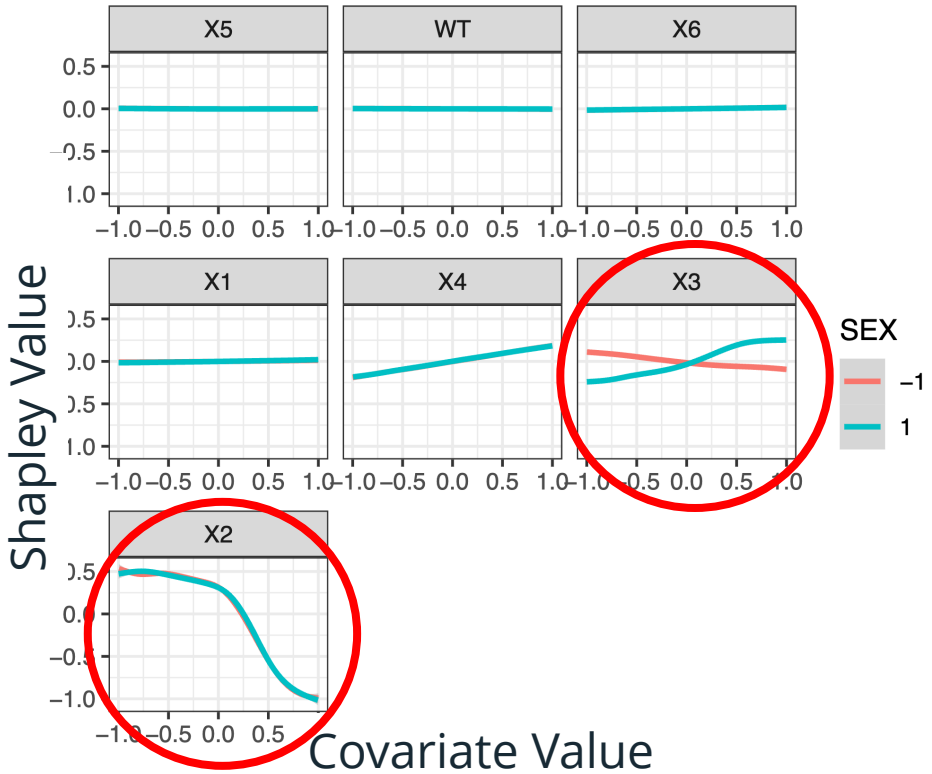


Wiens et. al. American Conference on Pharmacometrics 2022

SHAPLEY VALUES: COVARIATE RELATIONSHIPS

Non-linearity and Interactions





Wiens et. al. American Conference on Pharmacometrics 2022

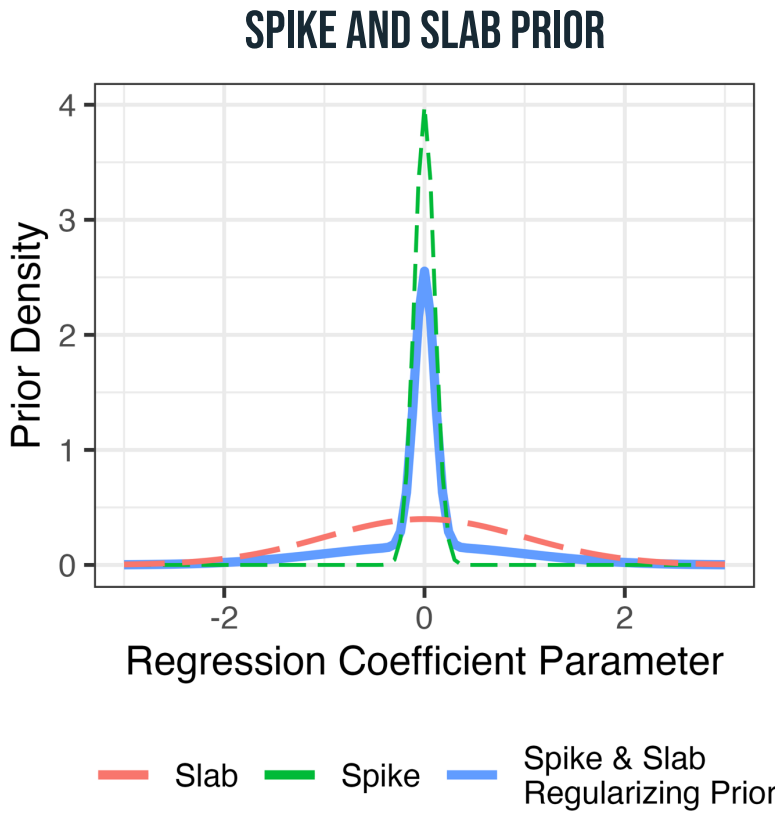
Wiens, M; Husband, H; Reilly, B; Callisto, S. American Conference on Pharmacometrics 2023

ASCPT 2024 ANNUAL MEETING

REGULARIZATION IN PARAMETRIC MODELS

What if we don't think we can estimate or need an ML model, but have lots of covariates?

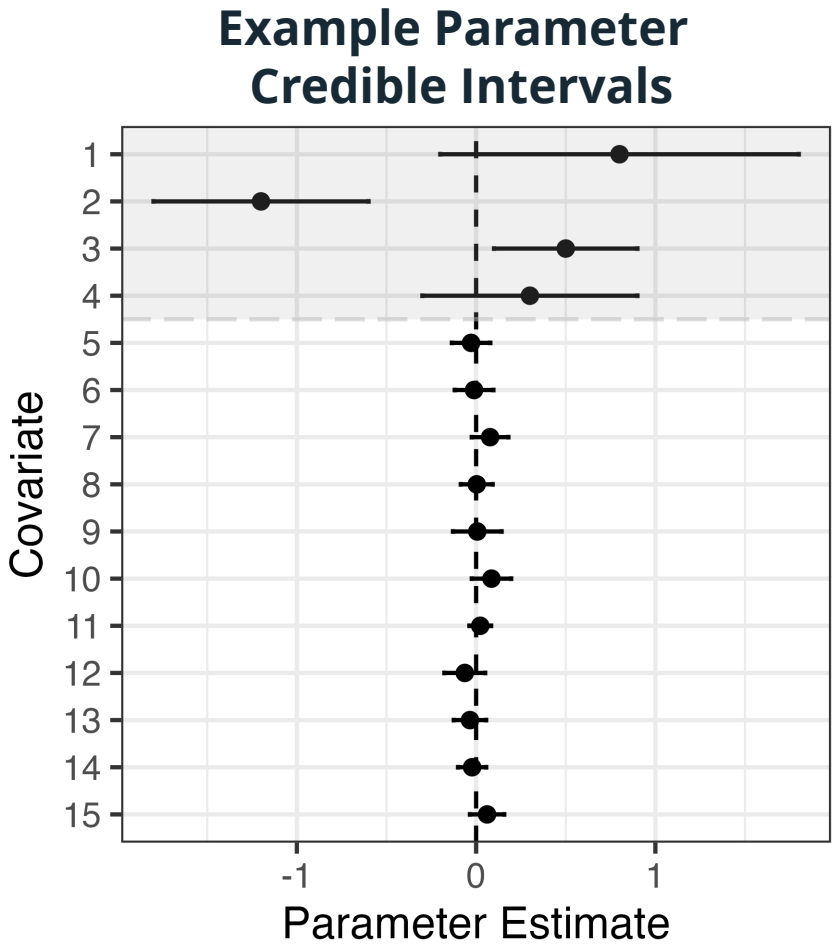
- Regularization for Bayesian Models
  - Shrink irrelevant effects to 0
  - Using informative priors
  - Priors are linked to real-world assumptions
- *Spike-and-Slab* or *Horseshoe* in exposure-response
- Straightforward implementation in Stan/brms



Garica, R and Rogers, J. Boston Pharmaceutical Symposium 2023

SPARSITY-INDUCING PRIORS FOR VARIABLE SELECTION: RESULTS

- Clear differences between important covariates and irrelevant covariates
- No need for correlation heuristics
  - Beware of confounding if using a lot of covariates
- Probabilistic inferences about variable selection from the model



CONCLUSIONS

- Shapley Values for interpreting ML models
  - More than just finding important covariates
  - Can be applied to complicated parametric models
- Opportunities and applications for regularizing parametric models
- Value in incorporating pieces of ML into other analyses
  - Bayesian non-parametrics
  - Probability weights in causal inference