ADS 503 - Applied Predictive Modeling (M3)

Summer 2024 - Week 3

Dave Hurst

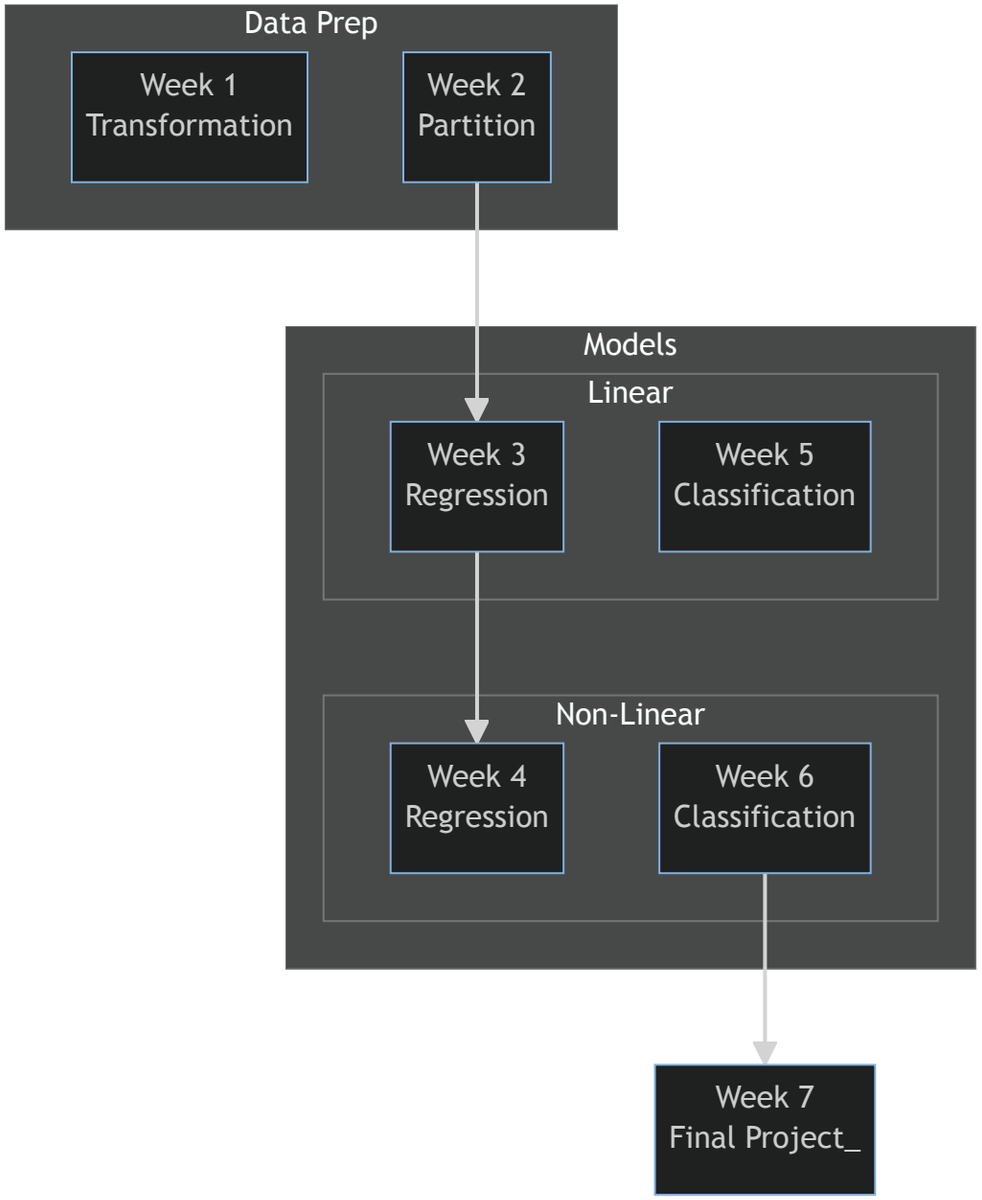
University of San Diego
SHILEY-MARCOS SCHOOL OF ENGINEERING

Start Recording!

University of San Diego
SHILEY-MARCOS SCHOOL OF ENGINEERING

Agenda

- Course Map
- System resources and Caching
- Assignment 2 Notes
- Assignment 3 Tips
- QA



University of San Diego
SHILEY-MARCOS SCHOOL OF ENGINEERING

System resources and Caching

Assignment 2 Notes

- K-fold CV -vs- MultiFolds

2.3.b (5 points): Using tools described in this chapter, provide code for implementing your approach(es).

```
set.seed(503) # set seed for reproducibility
repeatedSplits <- createDataPartition(permeabilitydf$permeability, p = .80, times =

# ... or ...
set.seed(503) # set seed for reproducibility
cvSplits <- createMultiFolds(permeabilitydf$permeability, k = 10, times = 5)
```

Assignment 2 Notes

- K-fold CV -vs- MultiFolds

To implement in Assignment 3

```
# Linear Regression
set.seed(seed)  # for reproducibility
lm_model <- train(train_data, train_fat,
                  method = "lm",
                  preProcess = c("center", "scale"),
                  trControl = trainControl(method = "repeatedcv", repeats = 5)
                  )
```

or

```
                  ...
                  trControl = trainControl(method = "cv")
```

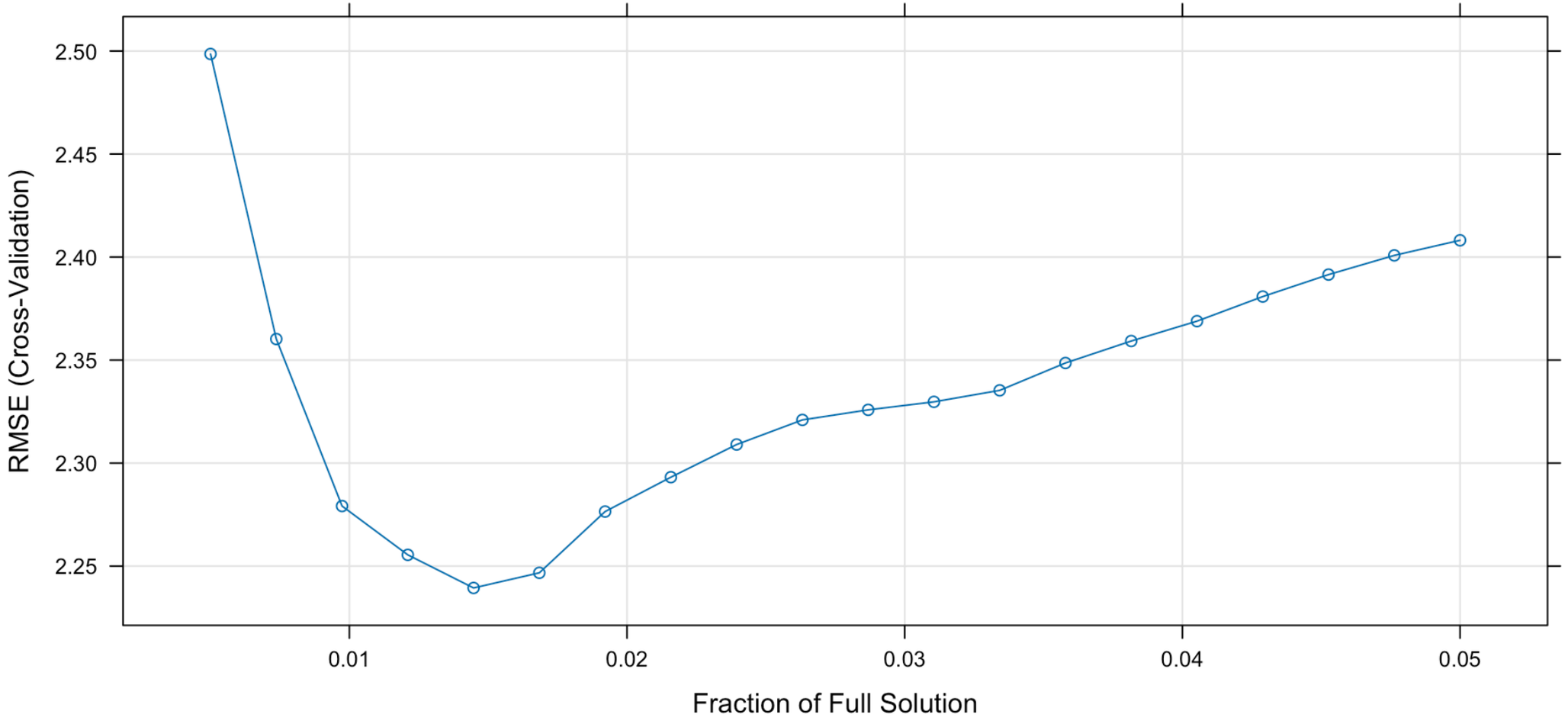
Assignment 3 Tips

- START EARLY!
- break 3.2.e into multiple chunks with `#| cache: true`
- on posit.cloud use `method = "cv"` **NOT** `method = "repeatedcv"`
- extracting results from objects in RStudio
  - fp_pls_model[[“bestTune”]][[“ncomp”]]
- set good limits for `tuneGrid`
  - use `plot(model)` to check limits
- ERRATA: `data(ChemicalManufacturingProcess)`
- extended PDF page limit = 25 (to allow for `plot(model)` – no data dumps)
- use `dotplot()` to compare models

```
1 lasso_model$bestTune
```

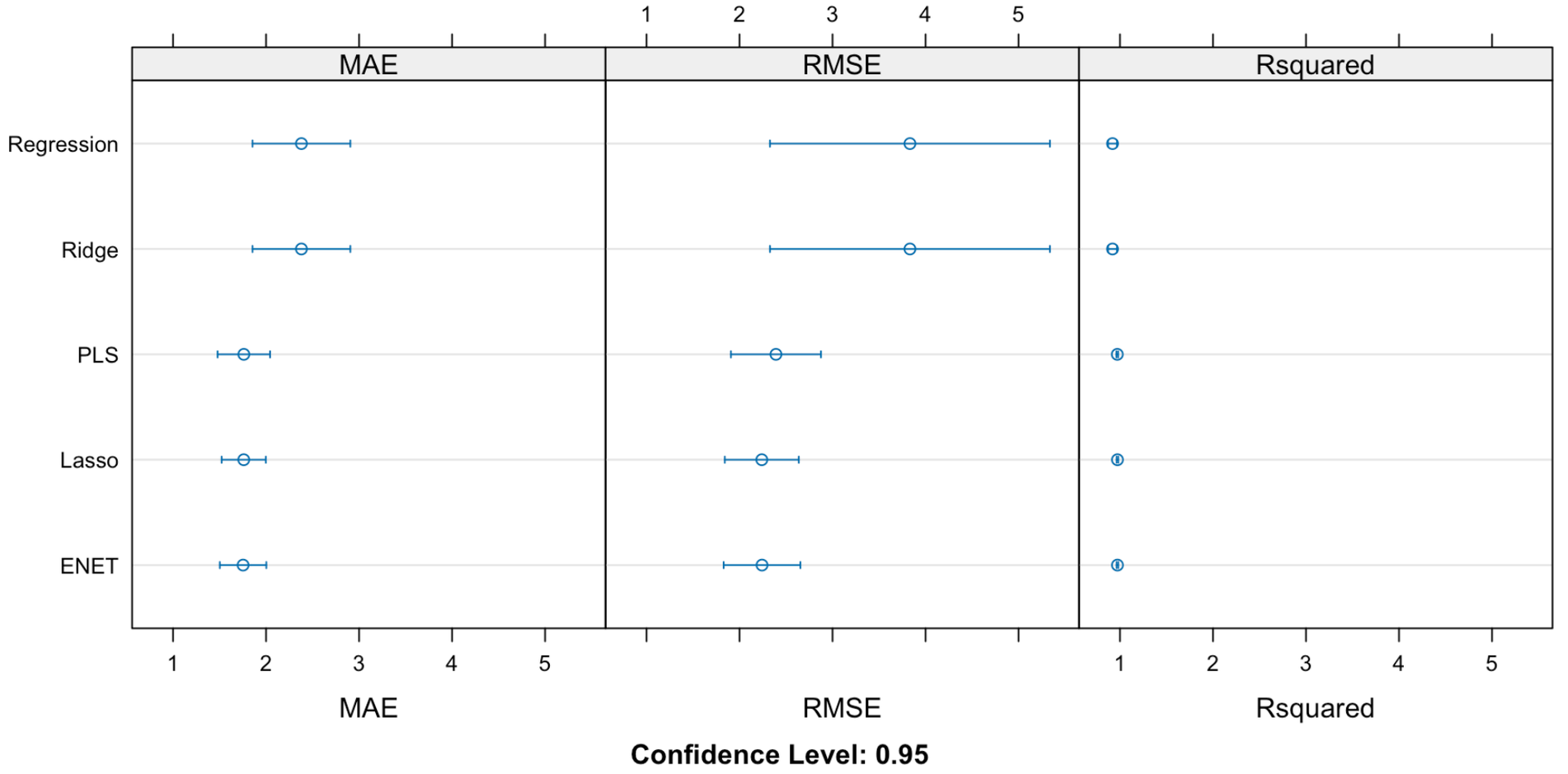
```
  fraction
5 0.01447368
```

```
1 plot(lasso_model)
```



University of San Diego
SHILEY-MARCOS SCHOOL OF ENGINEERING

```r
train_metrics <- resamples(list(
    Regression = lm_model,
    PLS = pls_model,
    Ridge = ridge_model,
    Lasso = lasso_model,
    ENET = enet_model))
dotplot(train_metrics)
```



Q&A

University of San Diego®
SHILEY-MARCOS SCHOOL OF ENGINEERING