DECISION PATTERNS

Caret

Practical Machine Learning (with R)

CARET

Caret

"Misc functions for training and plotting classification and regression models."

- Really:
  - Wraps 100's of modeling functions
  - Automates tediousness of model building
  - Manages a process
- Competitors:
  - Rattle : Graham Williams et al./Togaware.com
  - R Commander : Statistical workbench

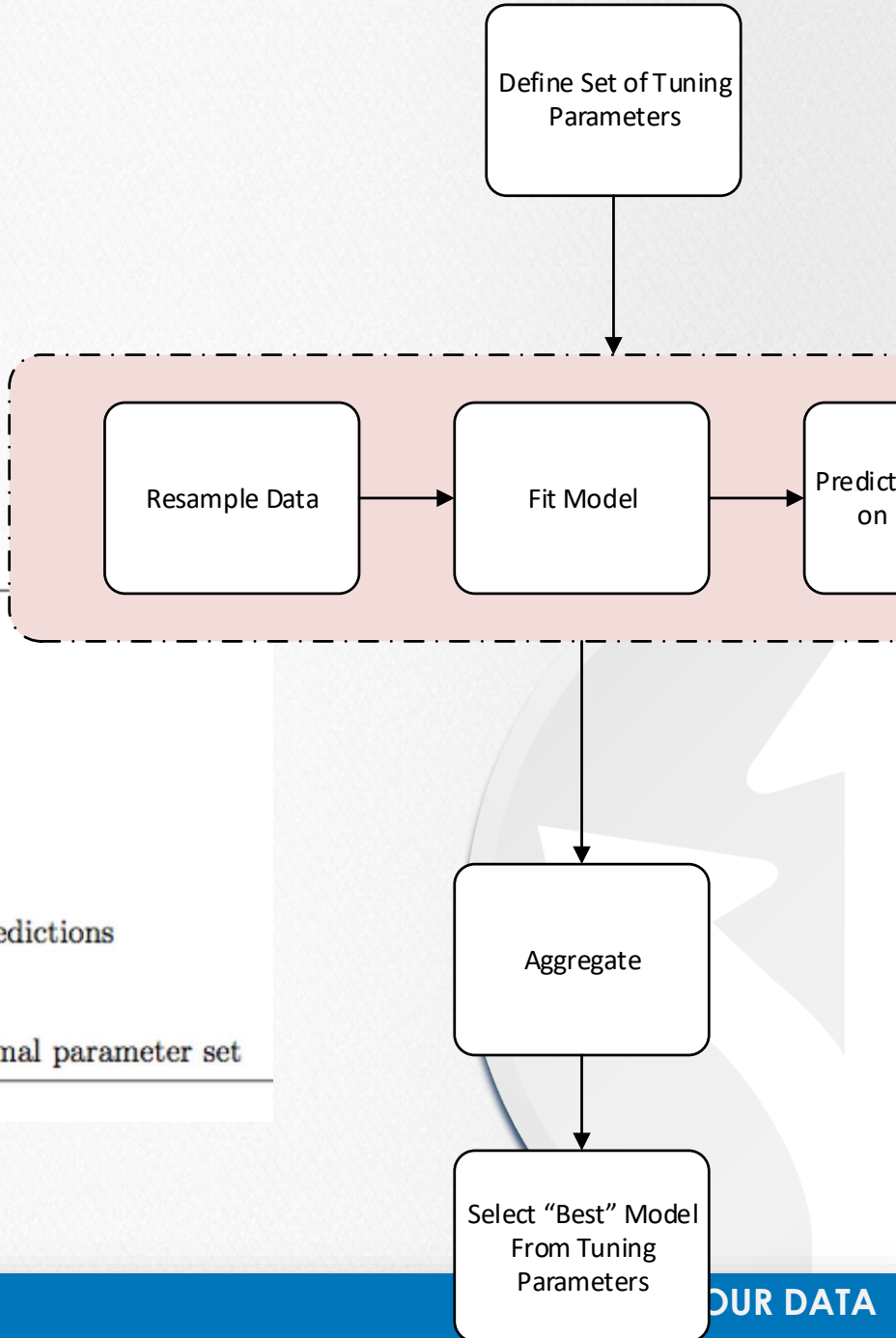
Caret Goals

Does a couple things:

- Preprocess data (transfroms, imputes)
- evaluate, using resampling, the effect of model tuning parameters on performance
- choose the "optimal" model across these parameters
- estimate model performance from a training set
- Variable Importance
- Aids feature selection

Process

```
Define sets of model parameter values to evaluate
for each parameter set do
    for each resampling iteration do
        Hold–out specific samples
        [Optional] Pre–process the data
        Fit the model on the remainder
        Predict the hold–out samples
    end
    Calculate the average performance across hold–out predictions
end
Determine the optimal parameter set
Fit the final model to all the training data using the optimal parameter set
```



Lots of Configurations

- Easy if you know what you are doing
- which `method`?

Caret Model List*

- Controlled mostly through
  - `train (tuneLength, tuneGrid)`
  - `trainControl – supplied to train`