Kinds of Ensembles

Tested on . . . dataset

David Nardi

June 11, 2024

MSc in AI, University of Florence

. . . dataset

Binary classification task

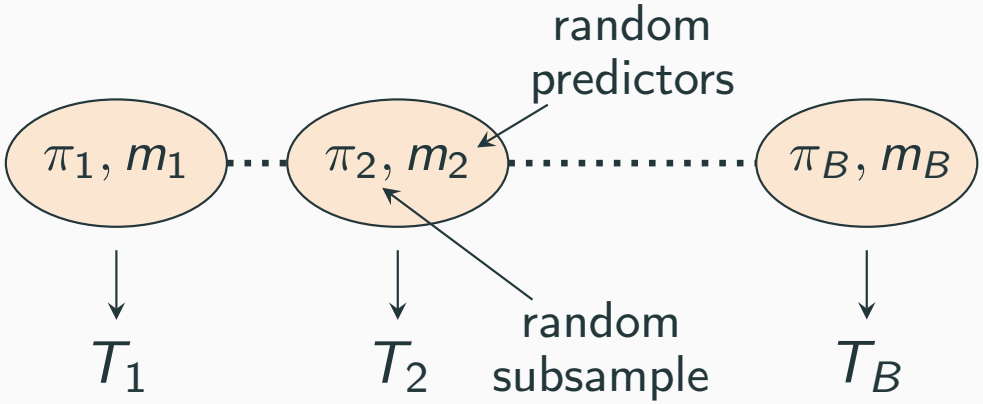
$\mathcal{Y} \in \{0, 1\}$

Class distribution: 0.49 – 0.51

Approaches

- Decision tree
- Random forest
- AdaBoost
- Ridge logistic regression
- Super learner

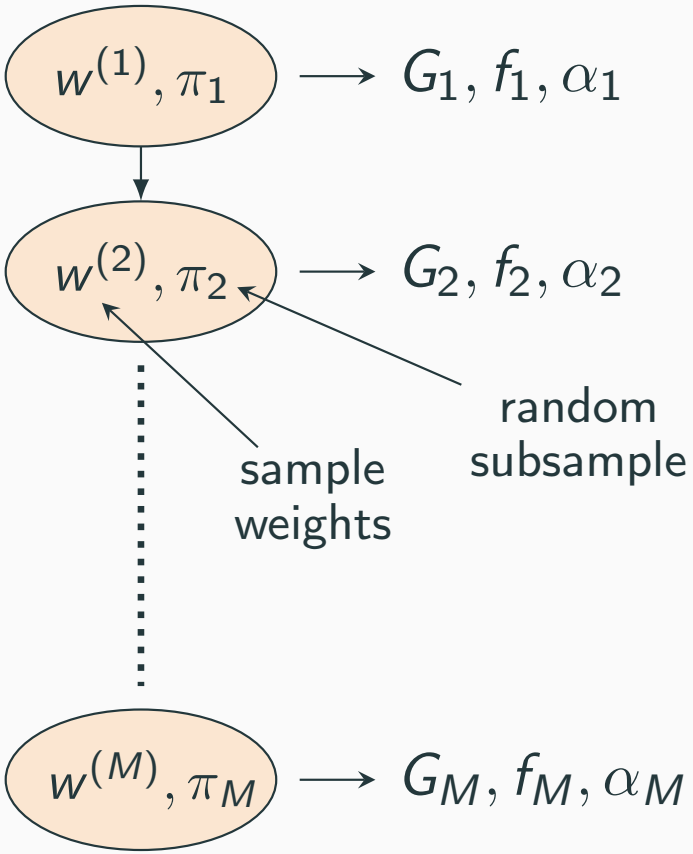
Random forest




```
randomForest(x, y,
  importance=TRUE,
  ntree=B.cross ← B*
  )
```

$$G(x) = \arg\max_k \sum_{b=1}^{B} \mathbb{I}(T_b(x) = k)$$

AdaBoost algorithm



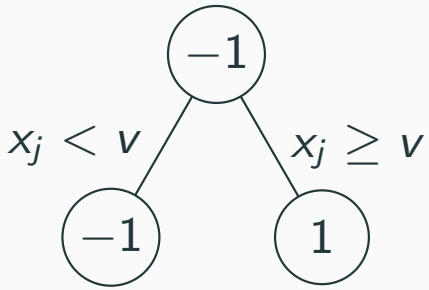
$$f_m(x) = f_{m-1}(x) + \lambda \alpha_m G_m(x)$$

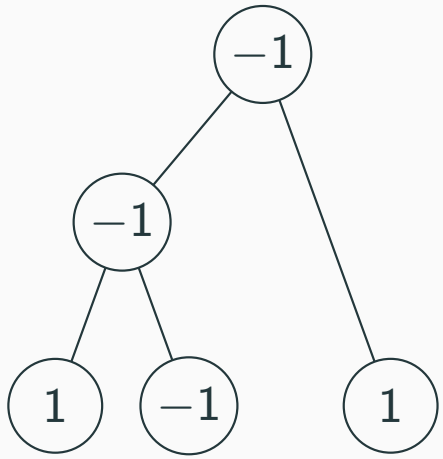
$$G(x) = \text{sign}(f_M(x))$$

$$L(y, f(x)) = \exp(-yf(x))$$

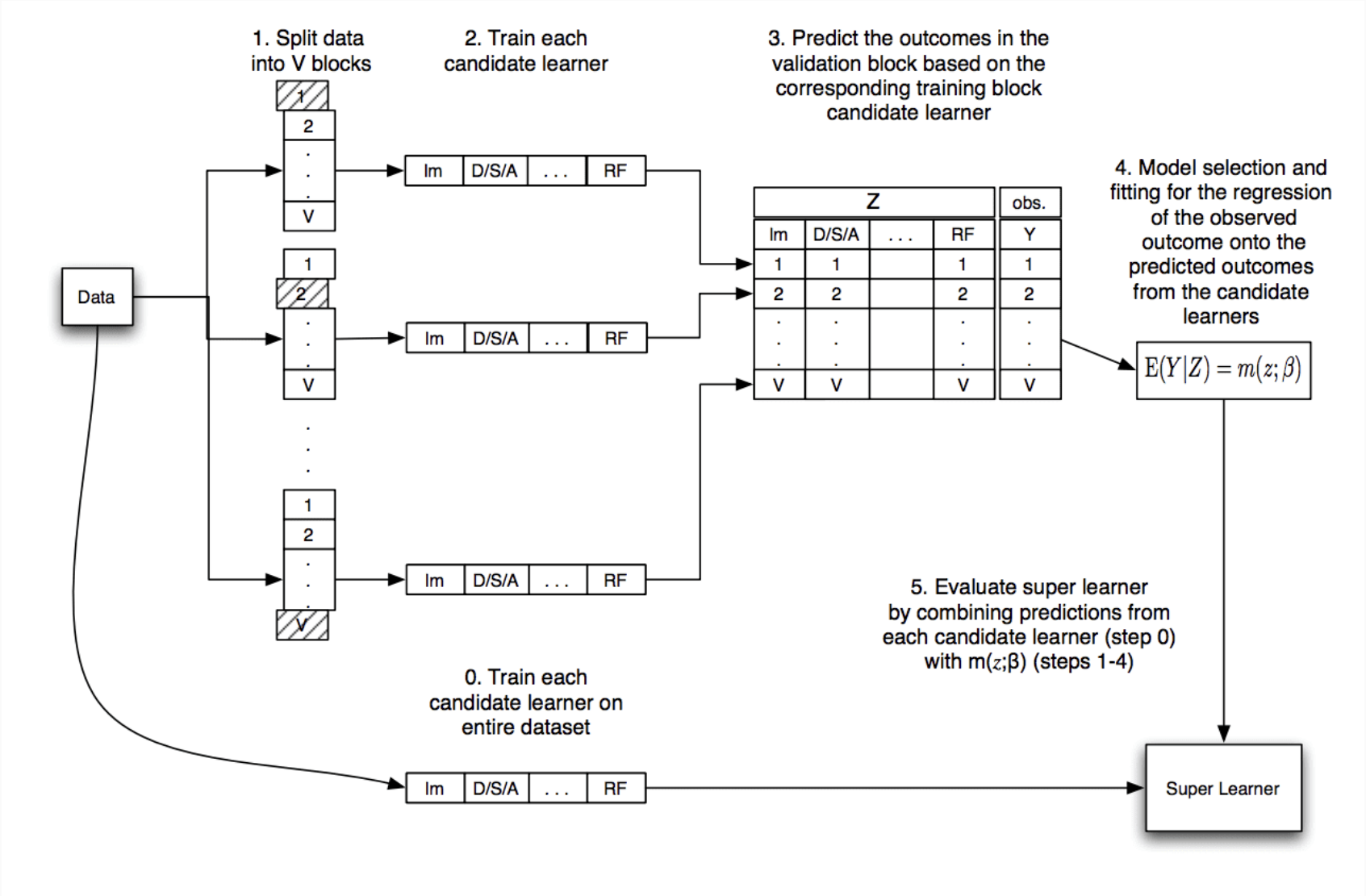
Encoding $\mathcal{Y} \in \{-1, 1\}$

```
ada::ada(x, y,
   loss="exponential",
   type="discrete",
   iter, ← M  nu=0.1, ← λ
   bag.frac=0.5, ← π
   control=stump ∨ depth2)
```

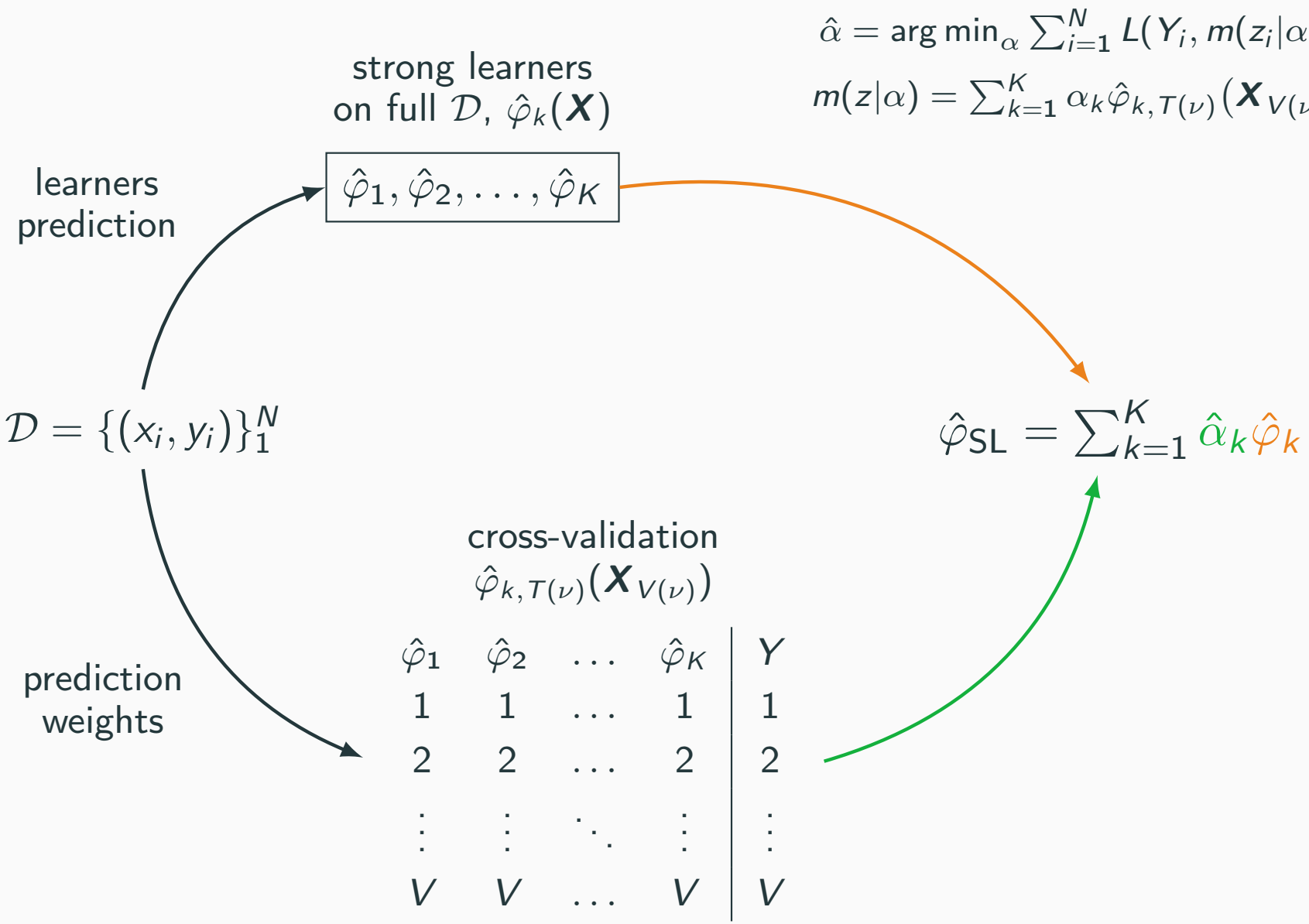




What's the Super Learner?



The Super Learner flow diagram



Super learner in practice

Package `SuperLearner`

Scores

Model	Train score	Test score
CART	0.0000	0.0000
Random forest	0.0000	0.0000
AdaBoost	0.0000	0.0000
Super learner	0.0000	0.0000

References i


E. C. Polley, and M. J. van der Laan
Super Learner in Prediction
U.C. Berkeley Division of Biostatistics Working Paper Series.
Working Paper 266, 2010