

Caret and zoon: machine learning, ecology and domain specific package systems

Tim C.D. Lucas
Malaria Atlas Project, BDI, Oxford

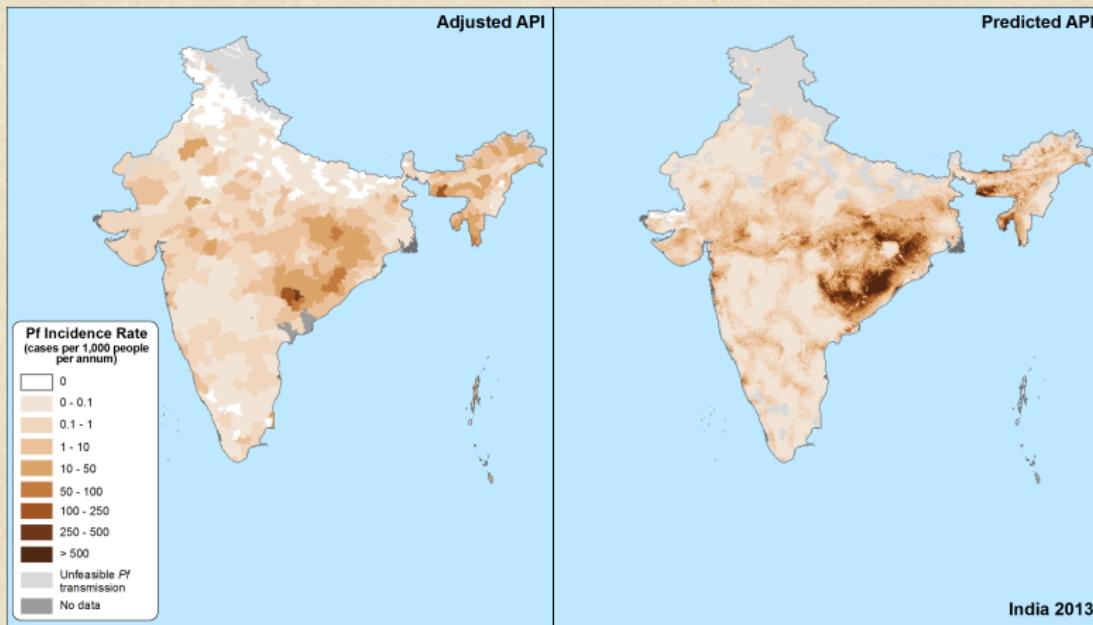


@timcdlucas @statsforbios
✉ timcdlucas@gmail.com

Who am I?

Malaria Atlas Project at BDI

Malaria, maps, geostatistics



Who am I?

R packages

Zoon

INLAutils

palettetown - my greatest ever achievement



Talk overview

caret

General package for machine learning.

Introduction to the package.

A domain specific package ecosystem?

zoon

General package for species distribution modelling.

What are SDMs?

Package overview.

Domain specific ecosystems

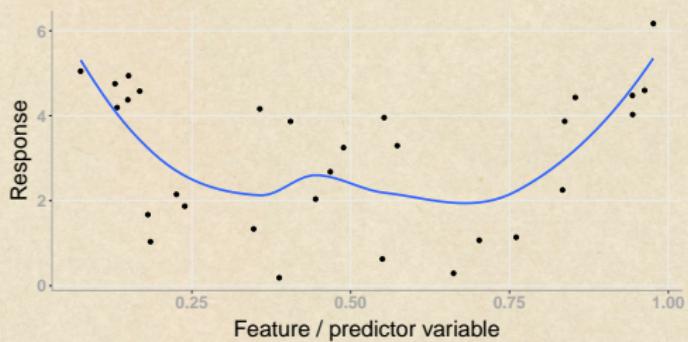
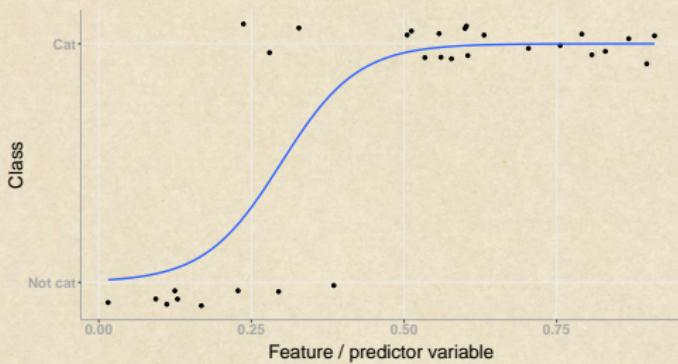
Other examples.

Are they a good thing?

caret

<https://topepo.github.io/caret/model-training-and-tuning.html>

What is machine learning?



Cross-validation

Hyperparameters

Hyperparameters

Number of PCA coordinates

Cut-offs for variable selection

$$x + x^2 + x^3 + x^4 + \dots$$

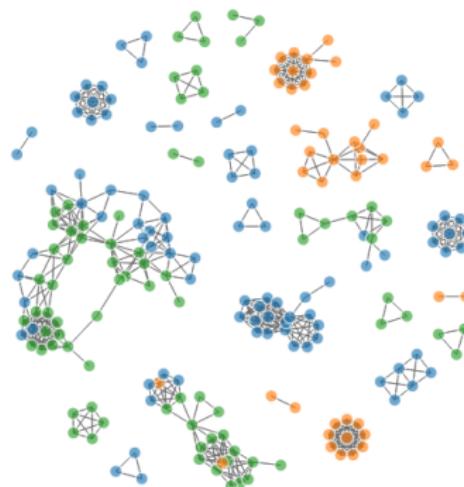
No free lunch

No such thing as a universal, 'best' machine learning model.

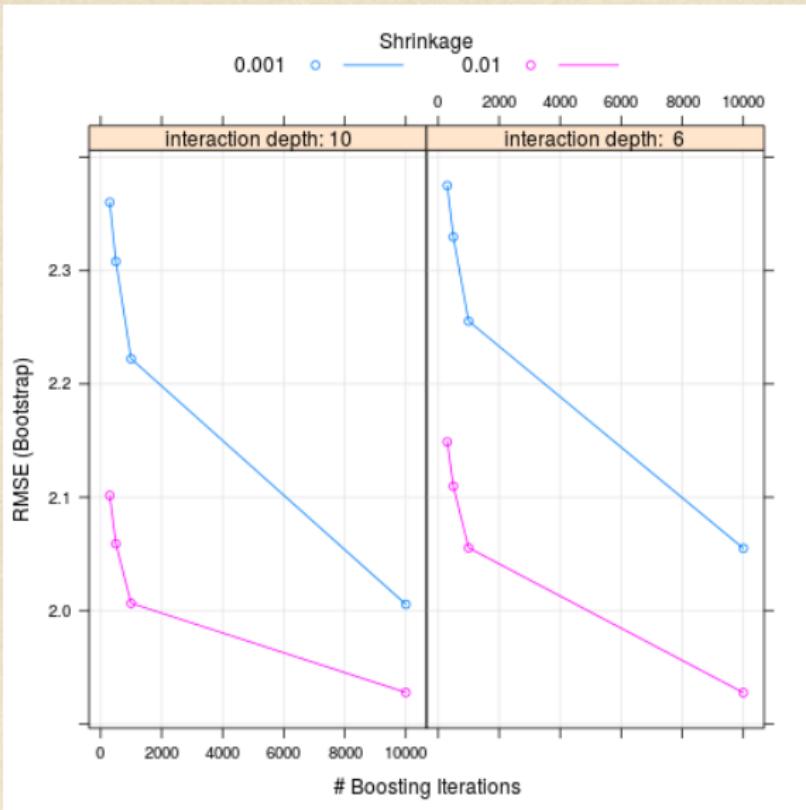
What does caret do?

8 Models Clustered by Tag Similarity

This page shows a network diagram of all the models that can be accessed by `train`. See the [Revolutions blog](#) for details about how this visualization was made (and [this page](#) has updated code using the `networkD3` package). In summary, the package annotates each model by a set of tags (e.g. "Bagging", "L1 Regularization" etc.). Using this information we can cluster models that are similar to each other.

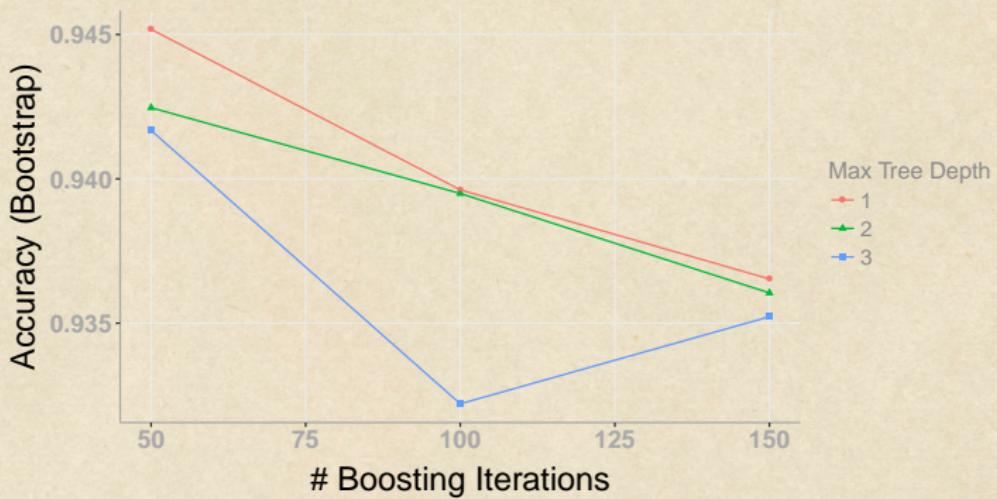


What does caret do?



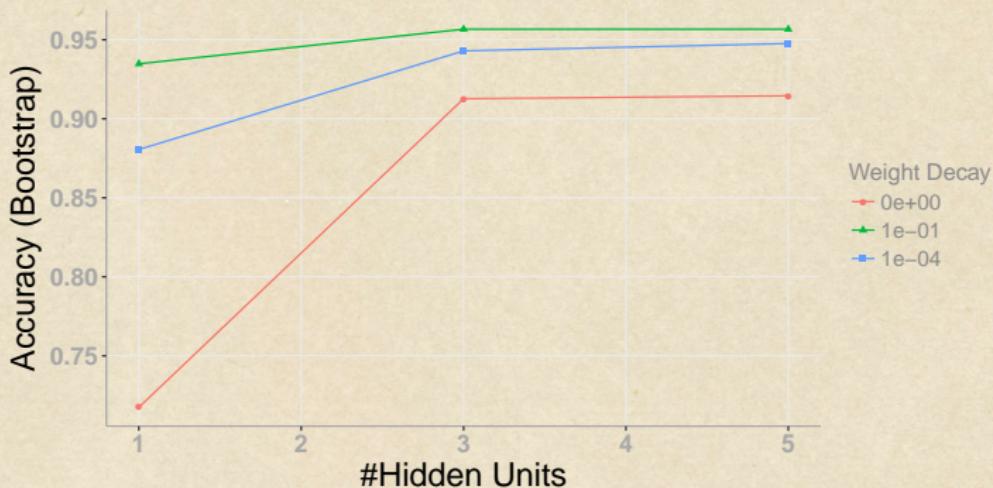
Training a model

```
m1 <- train(Species ~ .,  
             iris,  
             method = 'gbm')
```



Training a different model

```
m2 <- train(Species ~ .,  
             iris,  
             method = 'nnet')
```



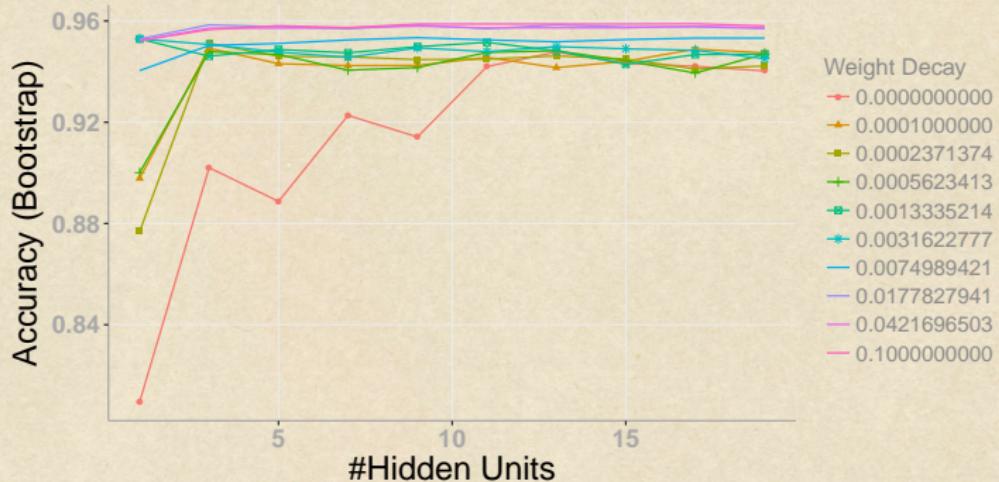
Controlling Crossvalidation

```
tr <- trainControl(method = 'cv', number = 5)

m3 <- train(Species ~ .,
             iris,
             trControl = tr,
             method = 'nnet')
```

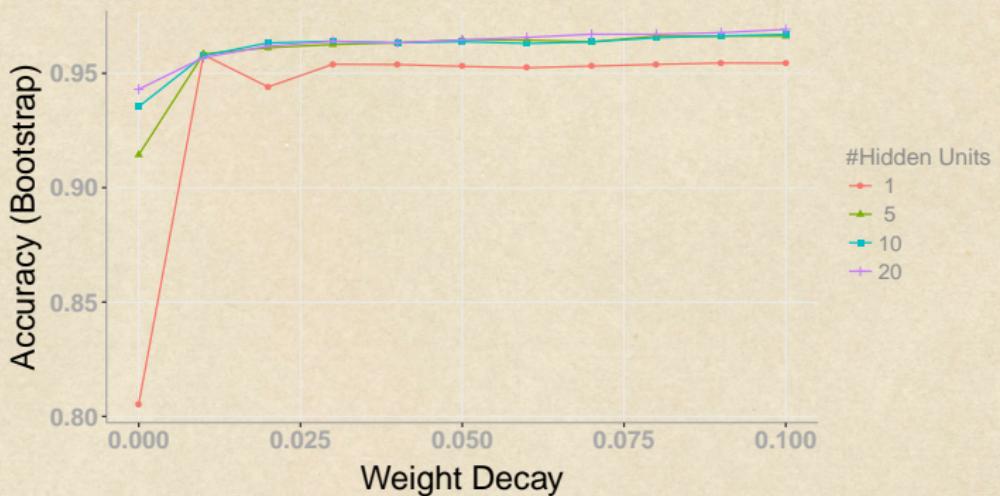
Try more hyperparameter values

```
m4 <- train(Species ~ .,  
             iris,  
             tuneLength = 10,  
             method = 'nnet')
```



Use chosen hyperparameter values

```
m5 <- train(Species ~ .,  
             iris,  
             tuneGrid = expand.grid(size=c(1,5,10,20),  
                                     decay=seq(0,0.1,0.01)),  
             method = 'nnet')
```



Contributions

Add your own models.

Share by github pull request.

But aim is for devs to keep package up to date.

Z O Ö N



Who develops zoon?

Tom August

Me

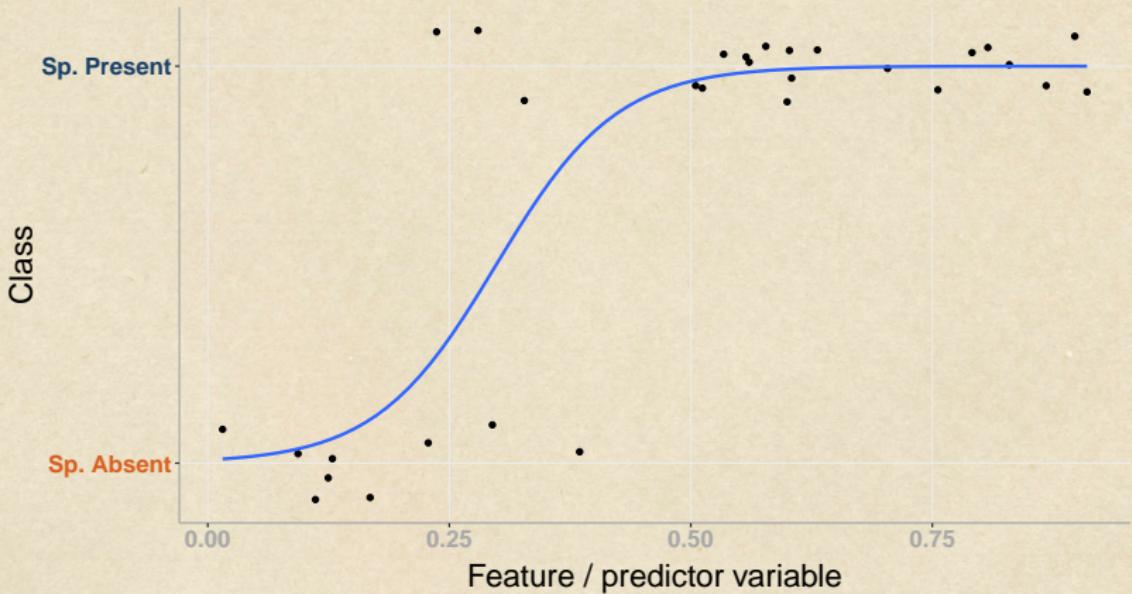
Nick Golding

Emiel van Loon

David Gavaghan

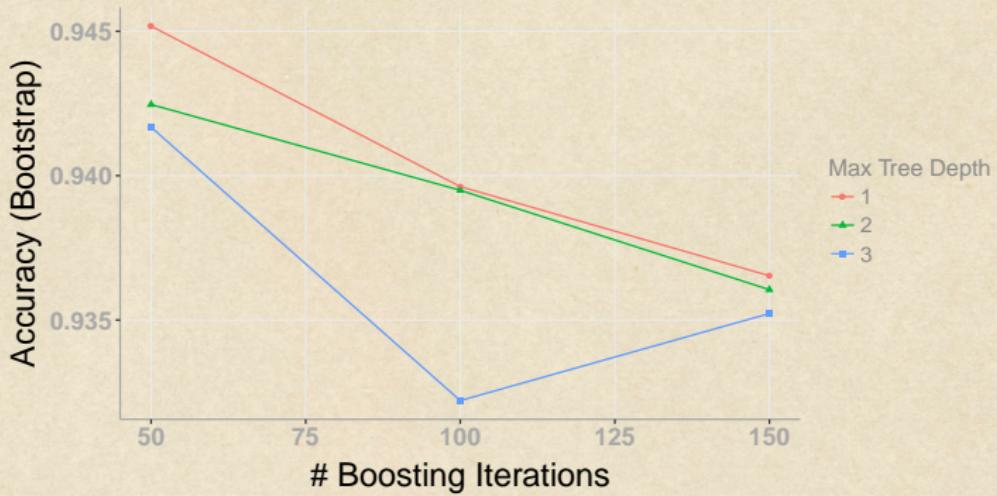
Greg McInerny

What does caret do?



A basic workflow

```
work1 <- workflow(  
  occurrence = UKAnophelesPlumbeus,  
  covariate = UKAir,  
  process = OneHundredBackground,  
  model = RandomForest,  
  output = PrintMap  
)
```



A different workflow

```
work2 <- workflow(  
  occurrence = UKAnophelesPlumbeus,  
  covariate = UKBioclim,  
  process = Background(n = 500),  
  model = RandomForest,  
  output = Appify)
```

A different model

```
work3 <- workflow(  
  occurrence = UKAnophelesPlumbeus,  
  covariate = UKAir,  
  process = OneHundredBackground,  
  model = MaxEnt,  
  output = PrintMap  
)
```



caret in zoon

```
work4 <- workflow(  
  occurrence = UKAnophelesPlumbeus,  
  covariate = UKAir,  
  process = OneHundredBackground,  
  model = MachineLearn(method = 'nnet',  
                        tuneLength = 8),  
  output = PrintMap  
)
```

Contributions

Add your own methods.

Share by web form or github.

Not the aim for devs to keep package up to date.

Package ecosystems

CRAN

zoon

dismo

Bioconductor

caret

Package ecosystems

CRAN

zoon

dismo

Bioconductor

caret

User contribution

Extendability

Any Questions ?

Tim C.D. Lucas