Multivariate Analysis II

HES 505 Fall 2022: Session 20

Matt Williamson

Objectives

By the end of today you should be able to:

- Articulate the differences between statisitical learning classifiers and logistic regression
- Describe several classification trees and their relationship to Random Forests
- Describe MaxEnt models for presence-only data

Revisiting Classification

Favorability in General

$$F(\mathbf{s}) = f(w_1 X_1(\mathbf{s}), w_2 X_2(\mathbf{s}), w_3 X_3(\mathbf{s}), \ldots, w_m X_m(\mathbf{s}))$$

- Logistic regression treats $f(x)$ as a (generalized) linear function
- Allows for multiple qualitative classes
- Ensures that estimates of $F(\mathbf{s})$ are [0,1]

Beyond Linearity

- Logistic (and other generalized linear models) are relatively interpretable
- Probability theory allows robust inference of effects
- Predictive power can be low
- Relaxing the linearity assumption can help

Classification Trees

- Use decision rules to segment the predictor space
- Series of consecutive decision rules form a ‘tree’
- Terminal nodes (leaves) are the outcome; internal nodes (branches) the splits

Classification Trees

- Divide the predictor space (R) into J non-overlapping regions
- Every observation in R_j gets the same prediction
- *Recursive binary splitting*
- Pruning and over-fitting

An Example

Inputs from the `dismo` package

An Example

The sample data

```r
head(pres.abs)
```

An Example

Building our dataframe

```r
pts.df <- terra::extract(pred.stack, vect(pres.abs), df=TRUE)
head(pts.df)
```

An Example

Building our dataframe

```r
pts.df[,2:7] <- scale(pts.df[,2:7])
summary(pts.df)
```

An example

- Fitting the classification tree

```
library(tree)
pts.df <- cbind(pts.df, pr
colnames(pts.df)[8] <- "y"
pts.df$y <- as.factor(ifel
tree.model <- tree(y ~ . ,
```

```
plot(tree.model)
text(tree.model, pretty=0)
```

An example

- Fitting the classification tree

```
summary(tree.model)
```

Benefits and drawbacks

Benefits

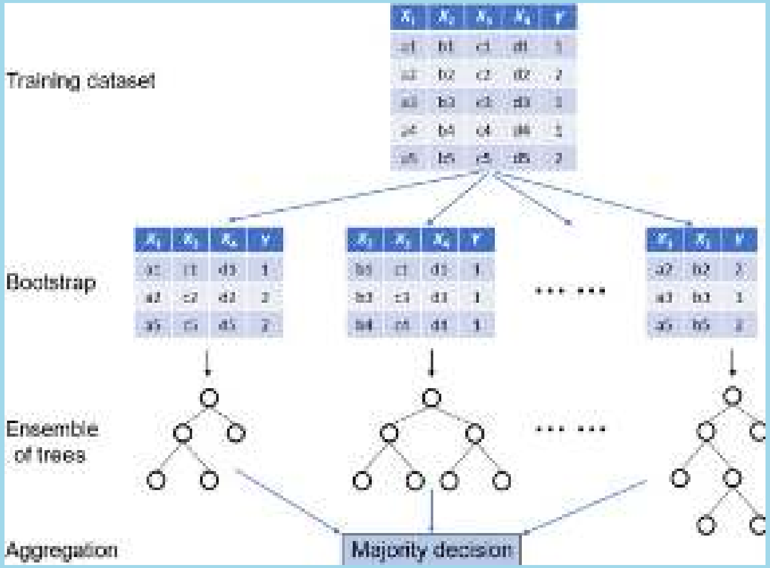
- Easy to explain
- Links to human decision-making
- Graphical displays
- Easy handling of qualitative predictors

Costs

- Lower predictive accuracy than other methods
- Not necessarily robust

Random Forests

- Grow 100(000s) of trees using bootstrapping
- Random sample of predictors considered at each split
- Avoids correlation amongst multiple predictions
- Average of trees improves overall outcome (usually)
- Lots of extensions



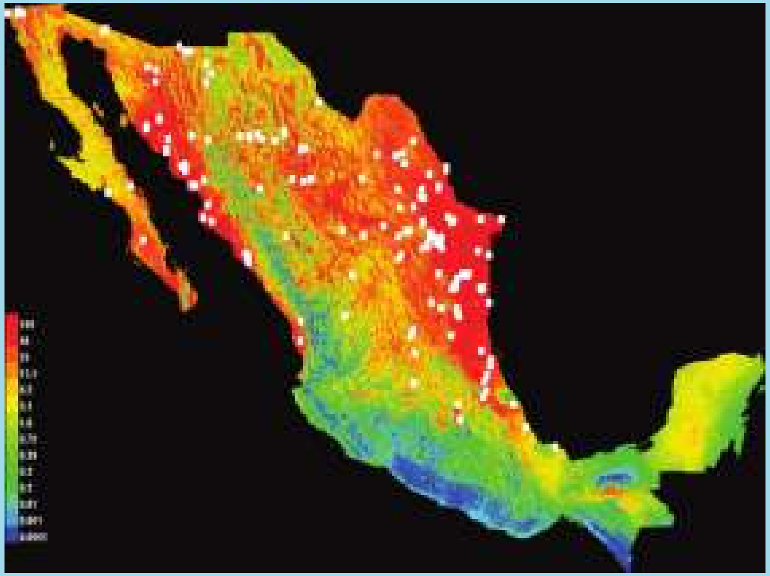
An example

- Fitting the Random Forest

```
library(randomForest)
class.model <- y ~ .
rf2 <- randomForest(class.
```

```
varImpPlot(rf2)
```

MaxEnt

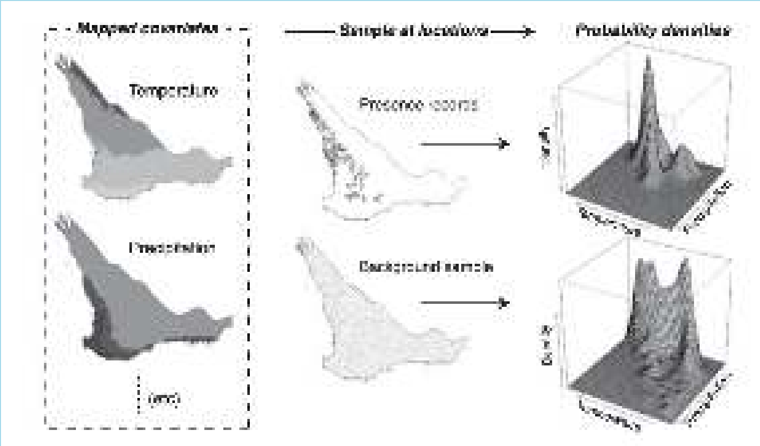


From Lentz et al. 2008

- Opportunistic collection of presences only
- Hypothesized predictors of occurrence are measured (or extracted) at each presence
- Background points (or pseudoabsences) generated for comparison

Maximum Entropy models

- MaxEnt (after the original software)
- Need *plausible* background points across the remainder of the study area
- Iterative fitting to maximize the distance between predictions generated by a spatially uniform model
- Tuning parameters to account for differences in sampling effort, placement of background points, etc
- Development of the model beyond the scope of this course, but see Elith et al. 2010



From Elith et al. 2010

Challenges with MaxEnt

- Not measuring *probability*, but relative likelihood of occurrence
- Sampling bias affects estimation (but can be mitigated using tuning parameters)
- Theoretical issues with background points and the intercept
- Recent developments relate MaxEnt (with cloglog links) to Inhomogenous Point Process models

Extensions

- Polynomial, splines, piece-wise regression
- Neural nets, Support Vector Machines, many many more