Consensus models weighted by AUC for multiple class responses

Andreas Dominik Cullmann

May 25, 2025

1 Introduction

This vignette shows how to build a consensus model for the `fgl` data (see [Venables and Ripley, 2002] or `?MASS::fgl`). We will follow the modelling process shown in [Marmion et al., 2009, Figure 1] restricting ourselves to only two 'Single-models': a classification tree and a multinomial log-linear model using package `rpart` and package `nnet`, respectively.

2 Example

After loading the data, we create random indices for training and evaluation sets (since the training and evaluation sets are complements with respect to the data set, two indices are redundant, but `fgl[ind_eval, ]` might be easier to read than `fgl[!ind_train, ]`).

```
> library(MASS)
> data(fgl)
> set.seed(100)
> ind_train <- sample(nrow(fgl), size = floor(nrow(fgl) * 0.7),
+                     replace = FALSE)
> ind_eval <- setdiff(seq(1:nrow(fgl)), ind_train)
```

Choosing `fgl$type` as response and all other variables as predictors, we calibrate a classification tree using `rpart::rpart` (cf. [Venables and Ripley, 2002, p. 264]):

```
> library(rpart)
> set.seed(123)
> fgl_rpart <- rpart(type ~ ., data = fgl[ind_train, TRUE],
+                    parms = list(split = "information"))
> newcp <- max(fgl_rpart$cptable[,"CP"] *
+              as.numeric(fgl_rpart$cptable[TRUE ,"xerror"] <
+                         sum(fgl_rpart$cptable[dim(fgl_rpart$cptable)[1],
+                             c("xerror","xstd")]))
+          ) + 1e-13
> fgl_rpart_pruned <- prune(fgl_rpart, cp = newcp)
```

and a multinomial log-linear model using `nnet::multinom`:

```
> library(nnet)
> fgl_multinom <- multinom(type ~ ., data = fgl[ind_train, TRUE],
+                          trace = FALSE)
```

We can now assess the model accuracy using either a confusion matrix of the classified predictions

```
> library(mda)
> confusion(predict(fgl_rpart_pruned, newdata = fgl[ind_eval, TRUE],
+                   type = "class"),
+           fgl[ind_eval, "type"])

         true
predicted WinF WinNF Veh Con Tabl Head
    WinF    18     7   3   0    0    0
    WinNF    3    13   3   0    0    1
    Veh      0     0   0   0    0    0
    Con      0     0   0   0    0    0
    Tabl     0     0   0   0    0    0
    Head     0     7   0   3    1    6

> confusion(predict(fgl_multinom, newdata = fgl[ind_eval, TRUE],
+                   type = "class"),
+           fgl[ind_eval, "type"])

         true
predicted WinF WinNF Veh Con Tabl Head
    WinF    12     2   4   0    0    1
    WinNF    6    19   0   0    0    0
    Veh      3     1   2   0    0    0
    Con      0     3   0   2    0    0
    Tabl     0     1   0   0    1    0
    Head     0     1   0   1    0    6
```

or a multiple class version of AUC using the raw predictions:

```
> library(HandTill2001)
> auc(multcap(response = fgl[ind_eval, "type"],
+             predicted = predict(fgl_rpart_pruned,
+                                 newdata = fgl[ind_eval, TRUE])))

[1] 0.7805996

> auc(multcap(response = fgl[ind_eval, "type"],
+             predicted = predict(fgl_multinom,
+                                 newdata = fgl[ind_eval, TRUE],
+                                 type = "probs")))

[1] 0.903246
```

To enhance predictive performance, we decide to use both models to build a consensus model. Furthermore, we want to use the 'weighted average' consensus method given by [Marmion et al., 2009, Eqn 1], which uses the pre-evaluated AUC of the models as weights. So we split the training set into two complementary subsets: 'inner training' and 'inner evaluation'.

```
> set.seed(100)
> ind_inner_train <- sample(ind_train,
+                           size = floor(length(ind_train) * 0.7),
+                           replace = FALSE)
> ind_inner_eval <- setdiff(ind_train, ind_inner_train)
```

We then refit our two models to the 'inner training' data:

```
> wa_fgl_multinom <- multinom(fgl_multinom, data = fgl[ind_inner_train, ],
+                             trace = FALSE)
> wa_fgl_rpart <- rpart(type ~ ., data = fgl[ind_inner_train, ],
+                       parms = list(split = "information")
+                       )
> newcp <- max(wa_fgl_rpart$cptable[,"CP"] *
+              as.numeric(wa_fgl_rpart$cptable[TRUE ,"xerror"] <
+                         sum(wa_fgl_rpart$cptable[dim(wa_fgl_rpart$cptable)[1],
+                             c("xerror","xstd")]))
+              ) + 1e-13
> wa_fgl_rpart_pruned <- prune(wa_fgl_rpart, cp = newcp)
```

and calculate pre-evaluation AUC (which we save in a `list`) using the 'inner evaluation' data and the refitted models:

```
> li <- list()
> li$rpart$auc <- auc(multcap(response = fgl[ind_inner_eval, "type"],
+                             predicted = predict(wa_fgl_rpart_pruned,
+                                                 newdata = fgl[ind_inner_eval,
+                                                               TRUE]
+                                                 )))
> li$mllm$auc <- auc(multcap(response = fgl[ind_inner_eval, "type"],
+                            predicted = predict(wa_fgl_multinom,
+                                                newdata = fgl[ind_inner_eval,
+                                                              TRUE],
+                                                type = "probs")))
```

We add the predictions using the models (the 'original' or 'Single' ones, not the refitted) for the evaluation set

```
> li$rpart$predictions <- predict(fgl_rpart_pruned,
+                                 newdata = fgl[ind_eval, TRUE])
> li$mllm$predictions <- predict(fgl_multinom,
+                                newdata = fgl[ind_eval, TRUE],
+                                type = "probs")
```

and obtain the consensus predictions as ([Marmion et al., 2009, Eqn 1])

```
> predicted <- Reduce('+', lapply(li, function(x)
+                                 x$auc * x$predictions)
+ ) / Reduce('+', sapply(li,"[", "auc"))
```

which perform (slightly) better than the predictions using the 'single models':

```
> auc(multcap(response = fgl[ind_eval, "type"],
+             predicted = predicted))
```

```
[1] 0.8958533
```

To classify the predicted probabilities, we choose the class with highest predicted probability (which.max gives the _first_ maximum):

```
> classes_predicted <-
+     factor(x = apply(predicted, 1, function(x)
+                      dimnames(predicted)[[2]][which.max(x)]),
+            levels = levels(fgl[ind_eval, "type"])
+            )
```

References

[Marmion et al., 2009] Marmion, M., Parviainen, M., Luoto, M., Heikkinen, R. K., and Thuiller, W. (2009). Evaluation of consensus methods in predictive species distribution modelling. *Diversity and Distributions*, 15(1):59–69.

[Venables and Ripley, 2002] Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. Springer, New York, fourth edition. ISBN 0-387-95457-0.