

Introduction to bagged trees

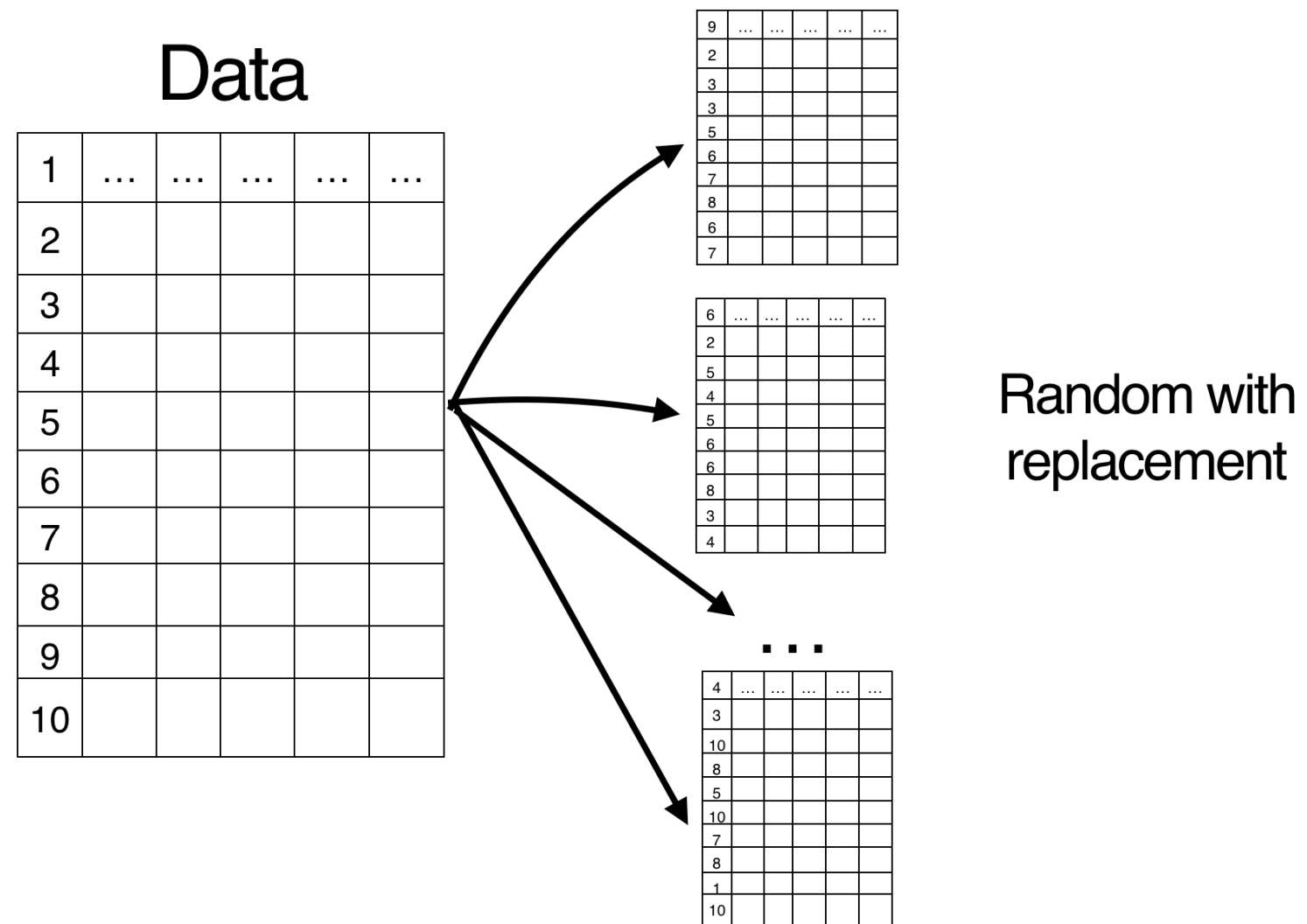
MACHINE LEARNING WITH TREE-BASED MODELS IN R



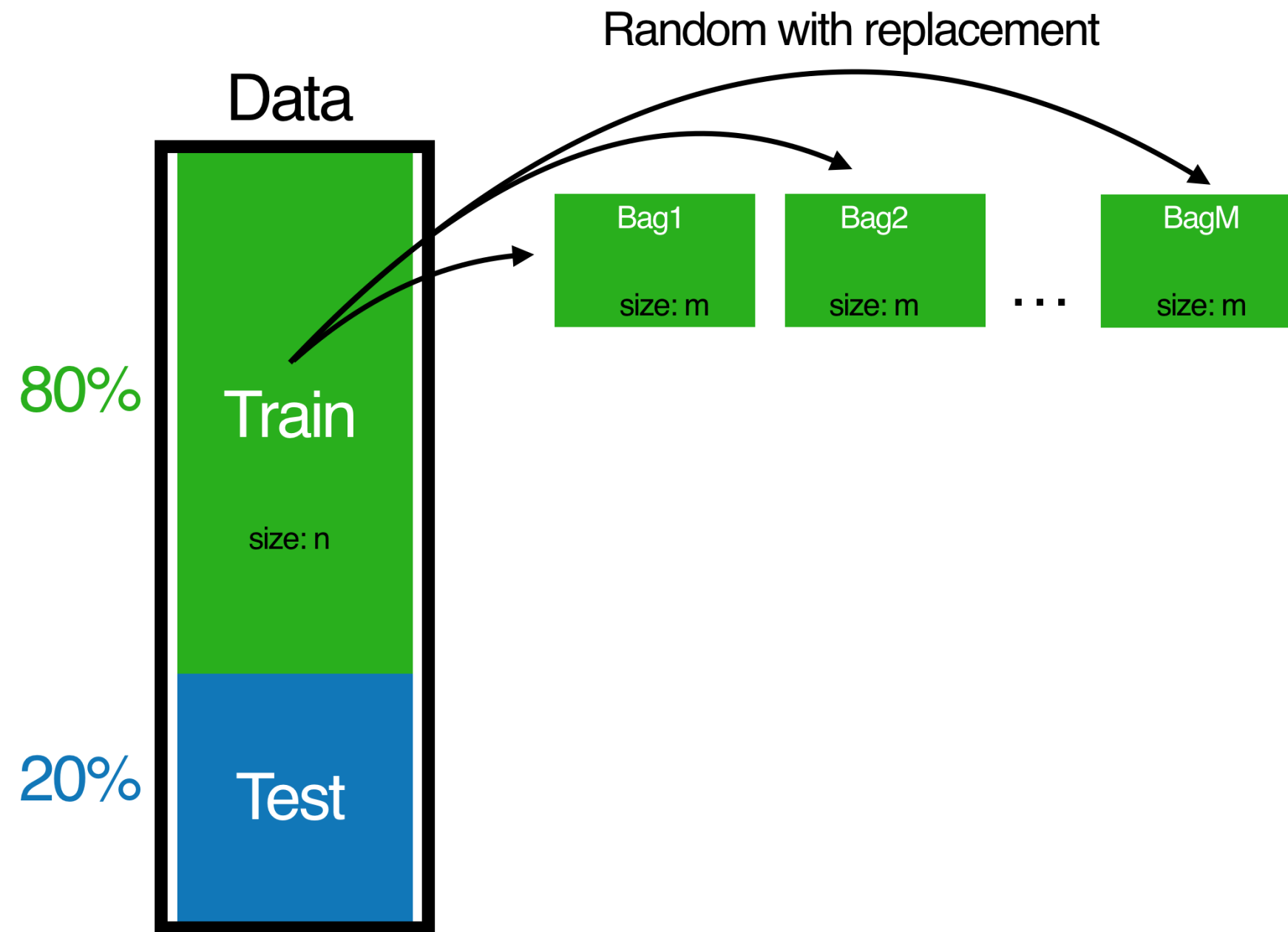
Gabriela de Queiroz
Instructor

Bagging

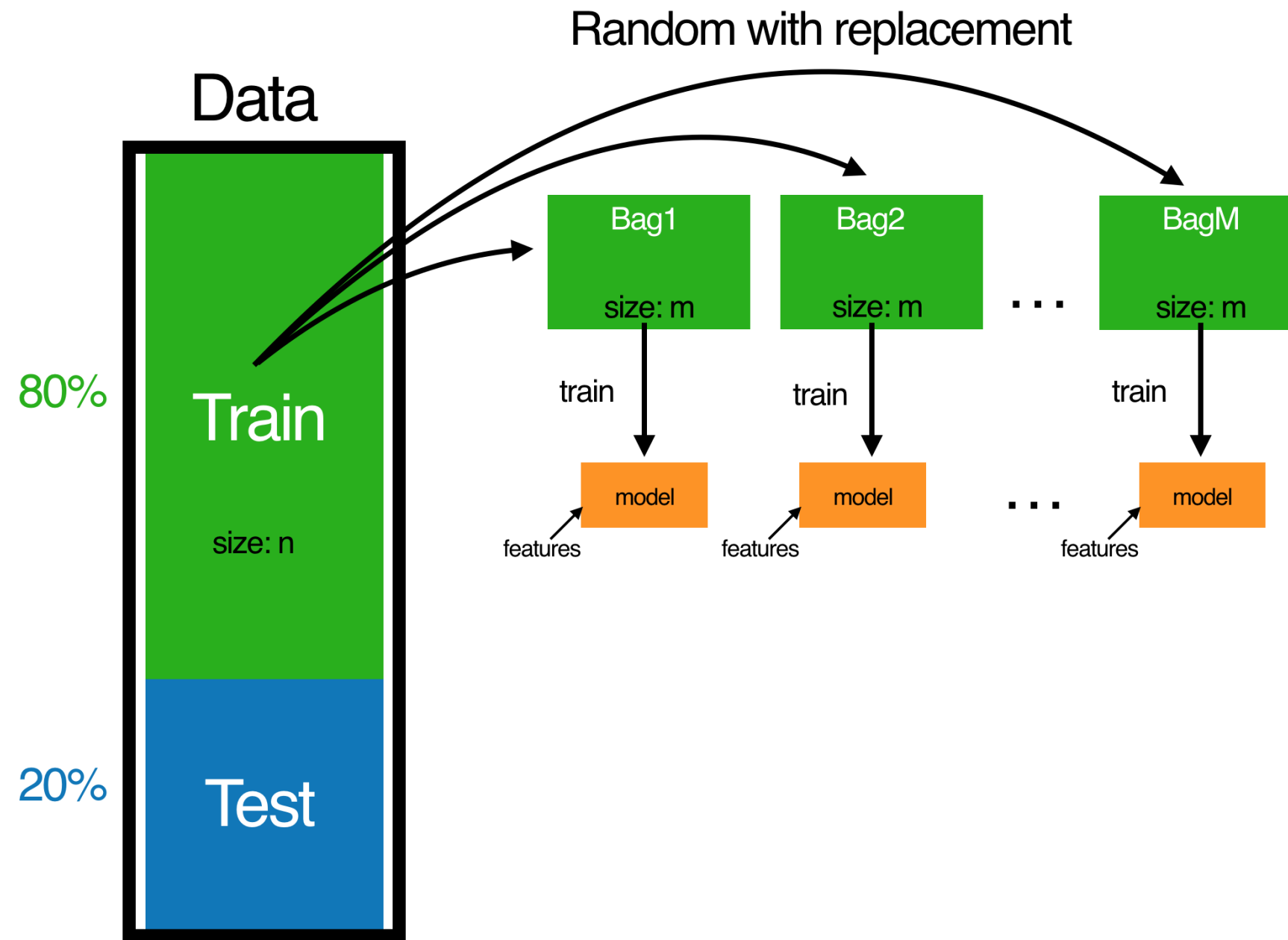
Bootstrap **AGG**regat**ING**



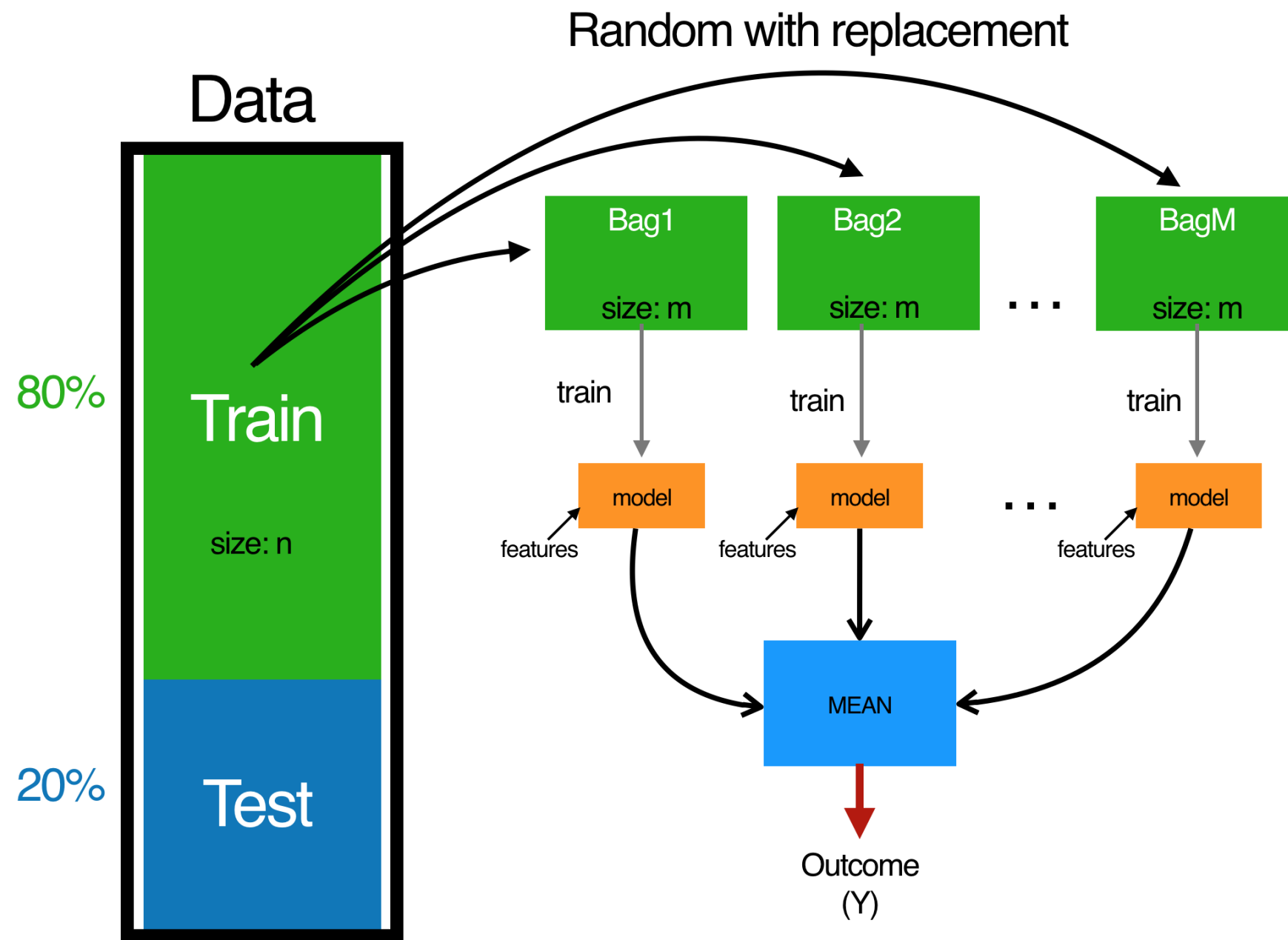
Step 1



Step 2



Bagging



Bagging in R

```
library(ipred)  
bagging(formula = response ~ ., data = dat)
```

Let's practice!

MACHINE LEARNING WITH TREE-BASED MODELS IN R

Evaluating the performance of bagged tree models

MACHINE LEARNING WITH TREE-BASED MODELS IN R



Gabriela de Queiroz
Instructor

Generate Predictions

```
class_predictions <- predict(object = rest_model_bag,  
                             newdata = restaurant_test,  
                             type = "class")
```

```
print(class_predictions)
```

```
Yes Yes Yes Yes No No Yes No ... Yes No No No Yes No  
Levels: No Yes
```

Confusion Matrix

```
confusionMatrix(data = class_pred,  
                 reference = restaurant_test$will_wait)
```

Confusion Matrix

Confusion Matrix and Statistics

	Reference	
Prediction	No	Yes
No	5	3
Yes	1	12

Accuracy : 0.8095

95% CI : (0.5809, 0.9455)

No Information Rate : 0.7143

P-Value [Acc > NIR] : 0.2402

...

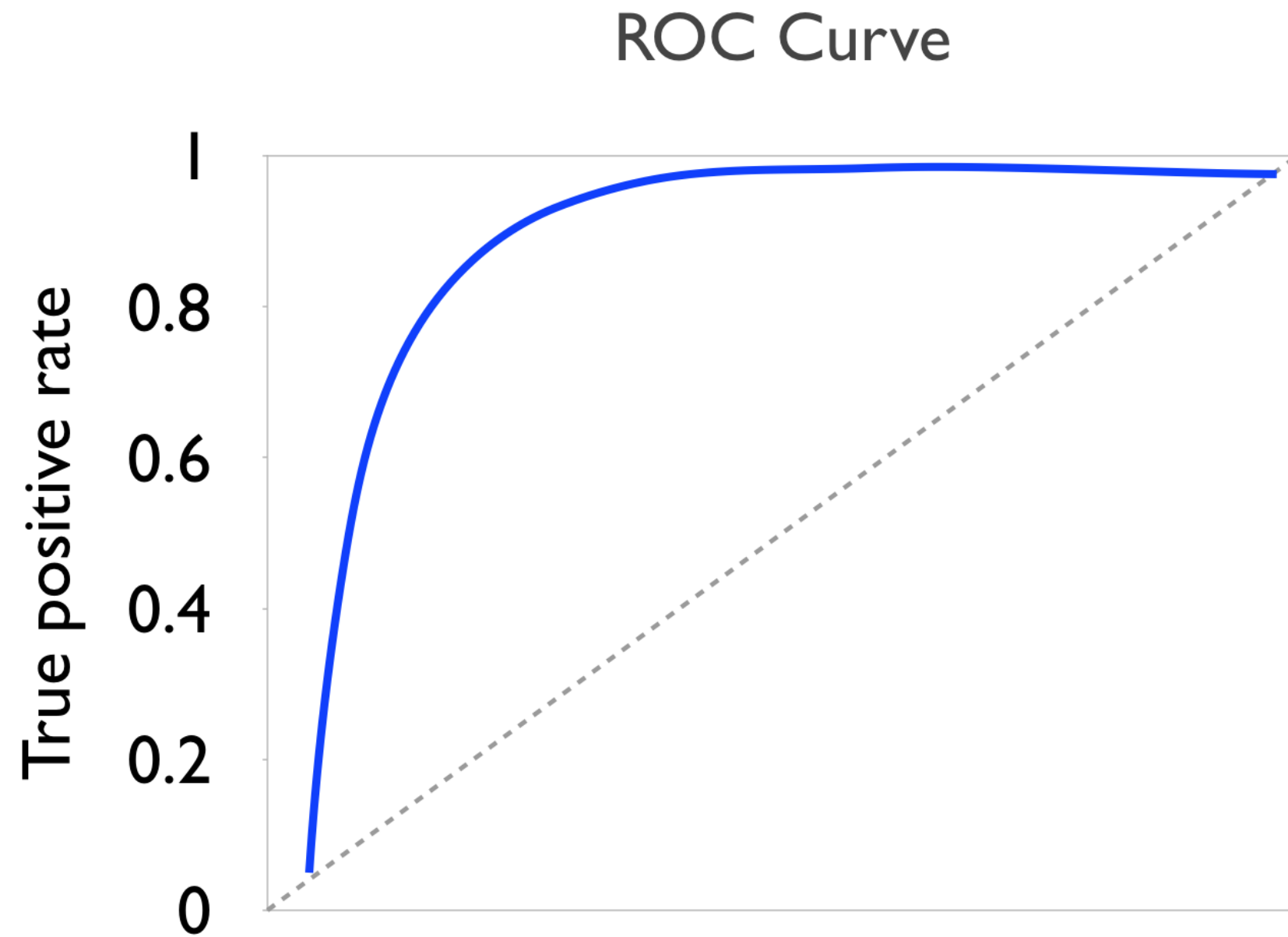
Sensitivity : 0.8333

Specificity : 0.8000

Pos Pred Value : 0.6250

Neg Pred Value : 0.9231

ROC Curve



AUC

```
library(Metrics)  
auc(actual, predicted)
```

```
.76765
```

Let's practice!

MACHINE LEARNING WITH TREE-BASED MODELS IN R

Cross-validation

MACHINE LEARNING WITH TREE-BASED MODELS IN R



Gabriela de Queiroz
Instructor

K-fold Cross-validation

- dataset size = 200 rows
- $k = 10$ (number of cross validation folds)

20	20	20	20	20
20	20	20	20	20

K-fold Cross-validation

20	20	20	20	20
20	20	20	20	20

- 10 estimates of test set AUC
- the average is the cross-validated estimate of AUC

Using caret for cross-validating models

```
library(caret)
```

- `train()`
- `trainControl()`

Training configuration

```
# Specify the training configuration
ctrl <- trainControl(method = "cv",
                     number = 5,
                     classProbs = TRUE,
                     summaryFunction = twoClassSummary)
```

Training configuration

```
# For reproducibility, set a seed
set.seed(1)
credit_model <- train(default ~ .,
                      data = credit_train,
                      method = "treebag",
                      metric = "ROC",
                      trControl = ctrl)
```

Let's practice!

MACHINE LEARNING WITH TREE-BASED MODELS IN R