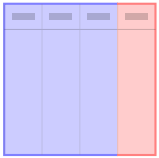
Applied Machine Learning

Machine Learning in R: MLR3 Resampling Strategies

Learning goals

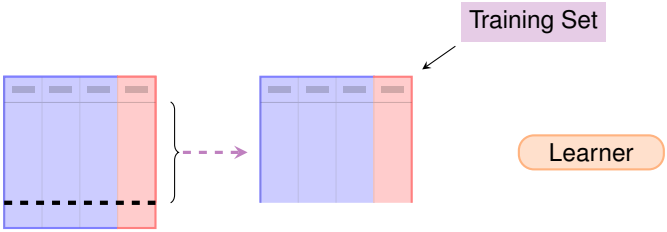
- Resampling in MLR3
- Working with ‘ResamplingResult‘ object

RESAMPLING

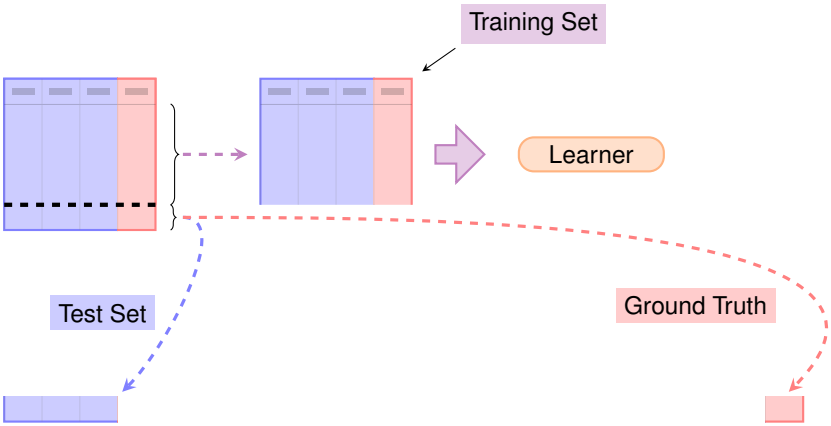


Learner

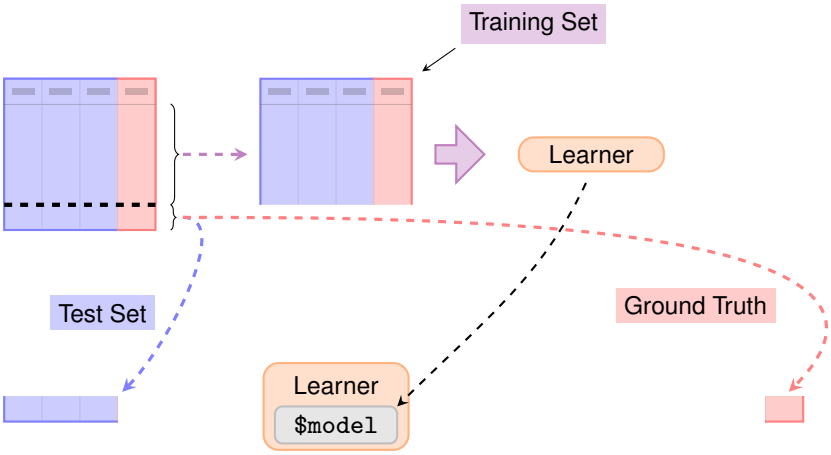
RESAMPLING



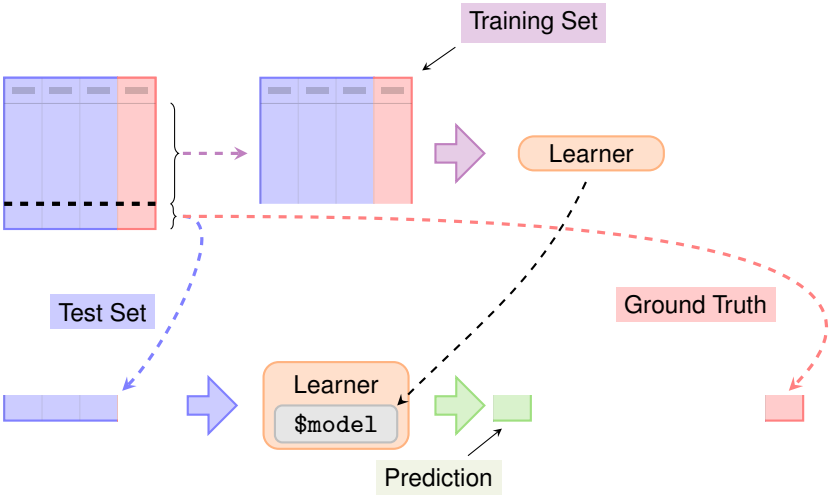
RESAMPLING



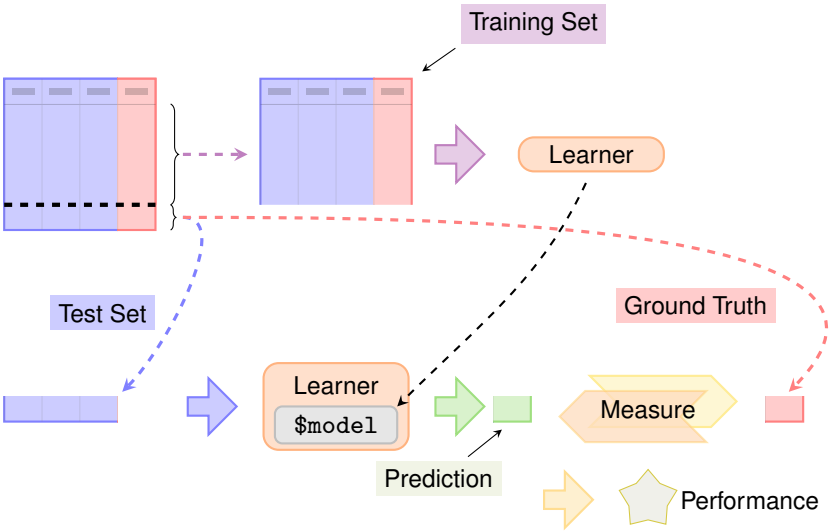
RESAMPLING



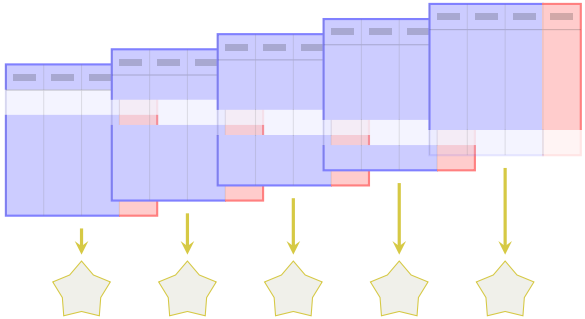
RESAMPLING


Training Set
Learner
Test Set
Ground Truth
Learner
$model
Prediction

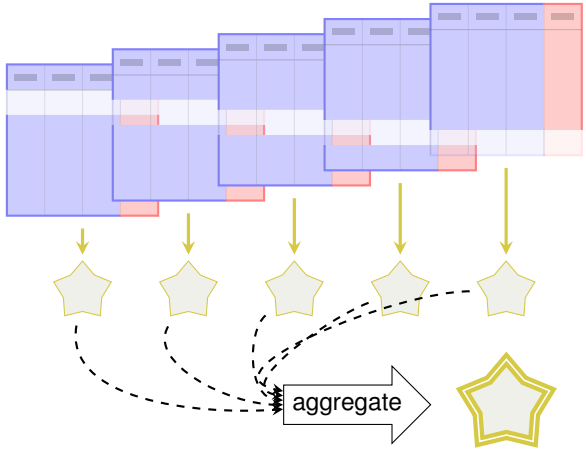
RESAMPLING


Training Set
Learner
Test Set
Ground Truth
Learner
$model
Measure
Prediction
Performance

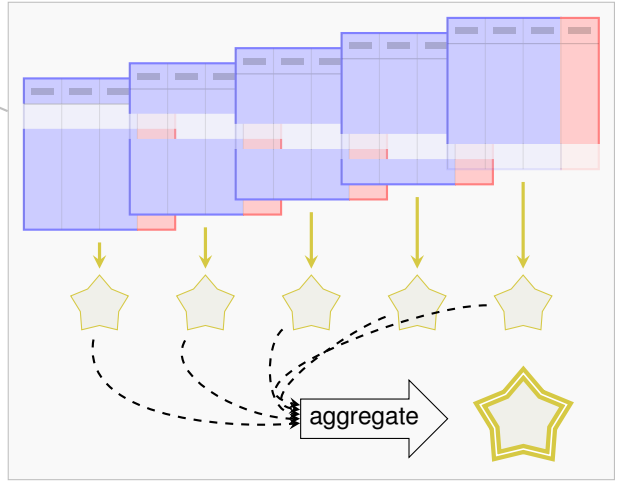
RESAMPLING



RESAMPLING



RESAMPLING



RESAMPLING

- Resample description: How to split the data

```
cv5 = rsmp("cv", folds = 5)
```

RESAMPLING

- Resample description: How to split the data

```
cv5 = rsmp("cv", folds = 5)
```

- Use the `resample()` function for resampling:

```
task = as_task_classif(x = iris, target = "Species", id = "iris")
learner = lrn("classif.rpart")
rr = resample(task, learner, cv5)
```

RESAMPLING

- Resample description: How to split the data

```
cv5 = rsmp("cv", folds = 5)
```

- Use the `resample()` function for resampling:

```
task = as_task_classif(x = iris, target = "Species", id = "iris")
learner = lrn("classif.rpart")
rr = resample(task, learner, cv5)
```

- We get a `ResamplingResult` object:

```
print(rr)
#> <ResampleResult> with 5 resampling iterations
#>  task_id    learner_id resampling_id iteration warnings errors
#>     iris classif.rpart            cv         1        0      0
#>     iris classif.rpart            cv         2        0      0
#>     iris classif.rpart            cv         3        0      0
#>     iris classif.rpart            cv         4        0      0
#>     iris classif.rpart            cv         5        0      0
```

RESAMPLING RESULTS

What exactly is a `ResamplingResult` object?
Remember `Prediction`:

- Get a table representation using `as.data.table()`

```
rr_table = as.data.table(rr)

print(rr_table)
#                   task                                learner
#                 <list>                                 <list>
# 1: <TaskClassif:iris> <LearnerClassifRpart:classif.rpart>
# 2: <TaskClassif:iris> <LearnerClassifRpart:classif.rpart>
# 3: <TaskClassif:iris> <LearnerClassifRpart:classif.rpart>
# 4: <TaskClassif:iris> <LearnerClassifRpart:classif.rpart>
# 5: <TaskClassif:iris> <LearnerClassifRpart:classif.rpart>
#        resampling iteration          prediction
#            <list>     <int>              <list>
# 1: <ResamplingCV>         1 <PredictionClassif>
# 2: <ResamplingCV>         2 <PredictionClassif>
# 3: <ResamplingCV>         3 <PredictionClassif>
# 4: <ResamplingCV>         4 <PredictionClassif>
# 5: <ResamplingCV>         5 <PredictionClassif>
```

- Active bindings and functions that make information easily accessible

RESAMPLING RESULTS

- Calculate performance:

```
rr$aggregate(msr("classif.ce"))
#> classif.ce
#>       0.06
```

RESAMPLING RESULTS

- Calculate performance:

```
rr$aggregate(msr("classif.ce"))
#> classif.ce
#>       0.06
```

- Get predictions

```
rr$prediction()
#> <PredictionClassif> for 150 observations:
#>     row_ids     truth  response
#>           2    setosa    setosa
#>           7    setosa    setosa
#>           9    setosa    setosa
#> ---
#>         141 virginica virginica
#>         148 virginica virginica
#>         149 virginica virginica
```

RESAMPLING

- Predictions of individual folds

```
predictions = rr$predictions()
predictions[[1]]
```

```
#> <PredictionClassif> for 30 observations:
#>     row_ids     truth   response
#>           2    setosa     setosa
#>           7    setosa     setosa
#>           9    setosa     setosa
#> ---
#>         135 virginica versicolor
#>         136 virginica  virginica
#>         142 virginica  virginica
```

RESAMPLING

- Predictions of individual folds

```
predictions = rr$predictions()
predictions[[1]]
```

```
#> <PredictionClassif> for 30 observations:
#>     row_ids     truth   response
#>           2    setosa     setosa
#>           7    setosa     setosa
#>           9    setosa     setosa
#> ---
#>         135 virginica versicolor
#>         136 virginica  virginica
#>         142 virginica  virginica
```

- Score of individual folds

```
scores = rr$score()
scores[1:3, c("iteration", "classif.ce")]
```

```
#>    iteration classif.ce
#>        <int>      <num>
#> 1:         1      0.067
#> 2:         2      0.100
#> 3:         3      0.067
```