Cross-validation demo

Simulated dataset are used for this demo. With 100 examples, and 5 explanatory variables:

```
# dataset creation
 set.seed(123)
 n <- 100 ; p <- 5
 X <- matrix(rnorm(n * p), n, p)
 y <- rnorm(n)
```

Define functions for calculating cross-validation error (MAPE and MAE):

- **MAPE**

```
# error measure 1: Mean Absolute Percentage Error - MAPE
eval_metric_mape <- function (preds, actual)
{
  res <- mean(abs(preds/actual-1))
  names(res) <- "MAPE"
  return(res)
}
```

- **MAE**

```
# error measure 2: Mean Absolute Error - MAE
eval_metric_mae <- function (preds, actual)
{
  res <- mean(abs(preds - actual))
  names(res) <- "MAE"
  return(res)
}
```

Linear model fitting, with RMSE, MAE and MAPE errors

`X` contains the explanatory variables.
`y` is the response.
`k` is the number of folds in k-fold cross-validation.
`repeats` is the number of repeats of the k-fold cross-validation procedure.

- **Defaut – Root Mean Squared Error – RMSE**

```
crossval::crossval_ml(x = X, y = y, k = 5, repeats = 3)
```

```
##
  |
  |                                                                      |   0%
  |
  |=============                                                         |  20%
  |
  |==========================                                            |  40%
  |
  |======================================                                |  60%
  |
  |================================================                      |  80%
  |
  |======================================================================| 100%
##    user  system elapsed
##   0.149   0.005   0.163
```

```
## $folds
##         repeat_1  repeat_2  repeat_3
## fold_1 0.8987732 0.9270326 0.7903096
## fold_2 0.8787553 0.8704522 1.2394063
## fold_3 1.0810407 0.7907543 1.3381991
## fold_4 1.0594537 1.1981031 0.7368007
## fold_5 0.7593157 0.8913229 0.7734180
##
## $mean
## [1] 0.9488758
##
## $sd
## [1] 0.1902999
##
## $median
## [1] 0.8913229
```

- **Mean Absolute Percentage Error – MAPE**

```
crossval::crossval_ml(x = X, y = y, k = 5, repeats = 3,
                      eval_metric = eval_metric_mape)
```

```
##
  |
  |                                                                      |   0%
  |
  |=============                                                         |  20%
  |
  |==========================                                            |  40%
  |
  |========================================                              |  60%
  |
  |======================================================                |  80%
  |
  |======================================================================| 100%
##    user  system elapsed
##   0.117   0.003   0.127
```

```
## $folds
##        repeat_1  repeat_2  repeat_3
## fold_1 1.486233 0.9517148 1.1181554
## fold_2 1.382454 1.1669799 1.0954839
## fold_3 1.267862 1.0583498 1.7768124
## fold_4 1.110386 1.1569593 1.3466701
## fold_5 1.242622 1.6604326 0.9615794
##
## $mean
## [1] 1.25218
##
## $sd
## [1] 0.2411539
##
## $median
## [1] 1.16698
```

- **Mean Absolute Error – MAE**

```
crossval::crossval_ml(x = X, y = y, k = 5, repeats = 3,
                      eval_metric = eval_metric_mae)
```

```
##
  |
  |                                                                      |   0%
  |
  |=============                                                         |  20%
  |
  |==========================                                            |  40%
  |
  |======================================                                |  60%
  |
  |================================================                      |  80%
  |
  |======================================================================| 100%
##    user  system elapsed
##   0.118   0.003   0.133

## $folds
##         repeat_1  repeat_2  repeat_3
## fold_1 0.7609698 0.6799802 0.6528781
## fold_2 0.7548409 0.7061494 0.9147533
## fold_3 0.8246641 0.5686014 1.0612401
## fold_4 0.7378648 0.9079500 0.5792025
## fold_5 0.6176459 0.7448324 0.6630864
##
## $mean
## [1] 0.7449773
##
## $sd
## [1] 0.1357212
##
## $median
## [1] 0.7378648…
```