Metrics

Thibault FUCHEZ

31/05/2021

```
library(ggplot2)
library(kableExtra)
library(dplyr)
```

```
##
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:kableExtra':
##
##     group_rows
```

```
## The following objects are masked from 'package:stats':
##
##     filter, lag
```

```
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

In order to compare the performance's models, we use differents "metrics". The reliability of our tools is higlhy dependant on the structure of our datas. In our case, the imbalanced sample of datas classes has to be take into account in order to find the metrics which allows to give a good evaluation of our models.

Confusion Matrix :

```
mc <- matrix(c("TP", "FP", "FN", "TN"), nrow = 2)
dimnames(mc) <- list(c("Positive", "Negative"), c("Positive", "Negative"))
as.data.frame(mc) %>% kable() %>% kable_styling() %>%
  pack_rows("True",1,2) %>% add_header_above(c(" " = 1 ,"Predicted" = 2))
```

```
# collapse_rows, group_rows, ... ???
```

True Positive rate : Recall, sensitivity, detection power

$$TP_{rate} = \frac{TP}{TP + FN}$$

	Predicted	
	Positive	Negative
True		
Positive	TP	FN
Negative	FP	TN

True Negative rate : Specificity

$$TN_{rate} = \frac{TN}{TN + FP}$$

False Positive rate : False alarm

$$FP_{rate} = \frac{FP}{TN + FP}$$

Global Error :

$$error = \frac{FP + FN}{TN + TP + FP + FN}$$

Positive prediction value : precision

$$PP_{value} = \frac{TP}{TP + FP}$$

Accuracy :

$$accuracy = 1 - error$$

dominance :

$$dominance = TP_{rate} - TN_{rate}$$

F-measure :

$$F_\beta = \frac{(1 + \beta^2) \times recall \times precision}{\beta^2 \times recall + precision}$$

Kappa :

AUC : Aera under the ROC curve

benefits(TP_{rate}) and costs (FP_{rate})