Bases de dados

Carregando R utilizado

```
source("R/GetDataSet.R")
source("R/GetDatasetdesc.R")
```

Existem diferentes bases disponiveis que podem ser usadas para o trabalho de classificação, reais e artificiais.

As bases reais foram retiradas do site da UCI https://archive.ics.uci.edu/ml/datasets.html e incluidos em arquivos csv. São bases de diferentes caracteristicas , numero de atributos e exemplares e diferentes niveis de desbalanceamento. As bases são as abaixo:

	Base	# Exemplares	Classe Positiva	% Classe Positiva	# Atributos	Link
1	Ecoli	336	pp	15%	7	https://archive.ics.uci.edu/ml/dataset
2	Glass	214	6	4%	9	https://archive.ics.uci.edu/ml/dataset
3	Haberman	306	2	26%	3	https://archive.ics.uci.edu/ml/dataset
4	Heart	303	1,2,3,4	4%	13	https://archive.ics.uci.edu/ml/dataset
5	Hepatitis	155	1	21%	19	https://archive.ics.uci.edu/ml/dataset
6	Iris	150	versicolor	33%	4	https://archive.ics.uci.edu/ml/dataset
7	Libra	360	1,2,3	20%	90	https://archive.ics.uci.edu/ml/dataset
8	Mamographic	961	Malign	46%	5	http://archive.ics.uci.edu/ml/datasets
9	Pima	768	1	35%	8	https://archive.ics.uci.edu/ml/machin
10	SPECTF-Heart	268	0	21%	44	https://archive.ics.uci.edu/ml/dataset
11	Wine	178	2	40%	13	https://archive.ics.uci.edu/ml/dataset
12	Wiscosin	699	malign	34%	9	https://archive.ics.uci.edu/ml/dataset

Table 1: Tabela com as bases utilizadas da UCI.

Temos diferentes bases identificadas pelo nome conforme abaixo:

```
ListClass <- c("Iris",
                "GlassLabel6",
                "ecoli",
                "haberman","wine","pima","libra1","libra123","vowel",
                "mamographic", "heart4", "heartnot0","wiscosin","newthyroid",
                "SPECTFheart","hepatitis")

ListClass
```

```
##  [1] "Iris"        "GlassLabel6" "ecoli"       "haberman"    "wine"
##  [6] "pima"        "libra1"      "libra123"    "vowel"       "mamographic"
## [11] "heart4"      "heartnot0"   "wiscosin"    "newthyroid"  "SPECTFheart"
## [16] "hepatitis"
```

As bases podem ser consultas usando a função R disponibilizada:

```
  DataSetName <- "GlassLabel6"
  ds <- GetDataSetMarcio(DataSetName)
```

```
## Parsed with column specification:
## cols(
##   RI = col_double(),
##   Na = col_double(),
##   Mg = col_double(),
##   Al = col_double(),
##   Si = col_double(),
##   K = col_double(),
##   Ca = col_double(),
##   Ba = col_double(),
##   Fe = col_double(),
##   Label = col_integer()
## )
```

```
  X <- ds$X
  Y <- ds$Y

  DataSetDesc <- GetDatasetdesc(DataSetName)

  head(X, 10)
```

```
##            RI    Na   Mg   Al    Si    K   Ca Ba   Fe
##  [1,] 1.52101 13.64 4.49 1.10 71.78 0.06 8.75  0 0.00
##  [2,] 1.51761 13.89 3.60 1.36 72.73 0.48 7.83  0 0.00
##  [3,] 1.51618 13.53 3.55 1.54 72.99 0.39 7.78  0 0.00
##  [4,] 1.51766 13.21 3.69 1.29 72.61 0.57 8.22  0 0.00
##  [5,] 1.51742 13.27 3.62 1.24 73.08 0.55 8.07  0 0.00
##  [6,] 1.51596 12.79 3.61 1.62 72.97 0.64 8.07  0 0.26
##  [7,] 1.51743 13.30 3.60 1.14 73.09 0.58 8.17  0 0.00
##  [8,] 1.51756 13.15 3.61 1.05 73.24 0.57 8.24  0 0.00
##  [9,] 1.51918 14.04 3.58 1.37 72.08 0.56 8.30  0 0.00
## [10,] 1.51755 13.00 3.60 1.36 72.99 0.57 8.40  0 0.11
```

```
    head(Y, 10)
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1
## Levels: 0 1
```

Também existem as bases artificiais são varias bases com diferentes niveis de sobreposição e desbalanceamento. As mesmas podem ser acessadas de forma semelhante:

```
ListMeanDifClass2 <- c(0:10)

ListMeanDifClass2 <- ListMeanDifClass2/2

ListnClassMin <- c(500,
409,
333,
269,
214,
167,
125,
88,
56,
26,
5
)

for (dif in ListMeanDifClass2)
{
    for (nmin in ListnClassMin)
  {
    DataSetName <-paste("overlap", dif, "_classmin", nmin, sep = "")
    print(DataSetName)
  }
}
```

```
## [1] "overlap0_classmin500"
## [1] "overlap0_classmin409"
## [1] "overlap0_classmin333"
## [1] "overlap0_classmin269"
## [1] "overlap0_classmin214"
## [1] "overlap0_classmin167"
## [1] "overlap0_classmin125"
## [1] "overlap0_classmin88"
## [1] "overlap0_classmin56"
## [1] "overlap0_classmin26"
## [1] "overlap0_classmin5"
## [1] "overlap0.5_classmin500"
## [1] "overlap0.5_classmin409"
## [1] "overlap0.5_classmin333"
## [1] "overlap0.5_classmin269"
## [1] "overlap0.5_classmin214"
## [1] "overlap0.5_classmin167"
## [1] "overlap0.5_classmin125"
## [1] "overlap0.5_classmin88"
## [1] "overlap0.5_classmin56"
## [1] "overlap0.5_classmin26"
## [1] "overlap0.5_classmin5"
## [1] "overlap1_classmin500"
## [1] "overlap1_classmin409"
## [1] "overlap1_classmin333"
## [1] "overlap1_classmin269"
## [1] "overlap1_classmin214"
## [1] "overlap1_classmin167"
## [1] "overlap1_classmin125"
## [1] "overlap1_classmin88"
## [1] "overlap1_classmin56"
## [1] "overlap1_classmin26"
## [1] "overlap1_classmin5"
## [1] "overlap1.5_classmin500"
## [1] "overlap1.5_classmin409"
## [1] "overlap1.5_classmin333"
## [1] "overlap1.5_classmin269"
## [1] "overlap1.5_classmin214"
## [1] "overlap1.5_classmin167"
## [1] "overlap1.5_classmin125"
## [1] "overlap1.5_classmin88"
## [1] "overlap1.5_classmin56"
## [1] "overlap1.5_classmin26"
## [1] "overlap1.5_classmin5"
## [1] "overlap2_classmin500"
## [1] "overlap2_classmin409"
## [1] "overlap2_classmin333"
## [1] "overlap2_classmin269"
## [1] "overlap2_classmin214"
## [1] "overlap2_classmin167"
## [1] "overlap2_classmin125"
## [1] "overlap2_classmin88"
## [1] "overlap2_classmin56"
## [1] "overlap2_classmin26"
## [1] "overlap2_classmin5"
## [1] "overlap2.5_classmin500"
## [1] "overlap2.5_classmin409"
## [1] "overlap2.5_classmin333"
## [1] "overlap2.5_classmin269"
## [1] "overlap2.5_classmin214"
## [1] "overlap2.5_classmin167"
## [1] "overlap2.5_classmin125"
## [1] "overlap2.5_classmin88"
## [1] "overlap2.5_classmin56"
## [1] "overlap2.5_classmin26"
## [1] "overlap2.5_classmin5"
## [1] "overlap3_classmin500"
## [1] "overlap3_classmin409"
## [1] "overlap3_classmin333"
## [1] "overlap3_classmin269"
## [1] "overlap3_classmin214"
## [1] "overlap3_classmin167"
## [1] "overlap3_classmin125"
## [1] "overlap3_classmin88"
## [1] "overlap3_classmin56"
## [1] "overlap3_classmin26"
## [1] "overlap3_classmin5"
## [1] "overlap3.5_classmin500"
## [1] "overlap3.5_classmin409"
## [1] "overlap3.5_classmin333"
## [1] "overlap3.5_classmin269"
## [1] "overlap3.5_classmin214"
## [1] "overlap3.5_classmin167"
## [1] "overlap3.5_classmin125"
## [1] "overlap3.5_classmin88"
## [1] "overlap3.5_classmin56"
## [1] "overlap3.5_classmin26"
## [1] "overlap3.5_classmin5"
## [1] "overlap4_classmin500"
## [1] "overlap4_classmin409"
## [1] "overlap4_classmin333"
## [1] "overlap4_classmin269"
## [1] "overlap4_classmin214"
## [1] "overlap4_classmin167"
## [1] "overlap4_classmin125"
## [1] "overlap4_classmin88"
## [1] "overlap4_classmin56"
## [1] "overlap4_classmin26"
## [1] "overlap4_classmin5"
## [1] "overlap4.5_classmin500"
## [1] "overlap4.5_classmin409"
## [1] "overlap4.5_classmin333"
## [1] "overlap4.5_classmin269"
## [1] "overlap4.5_classmin214"
## [1] "overlap4.5_classmin167"
## [1] "overlap4.5_classmin125"
## [1] "overlap4.5_classmin88"
## [1] "overlap4.5_classmin56"
## [1] "overlap4.5_classmin26"
## [1] "overlap4.5_classmin5"
## [1] "overlap5_classmin500"
## [1] "overlap5_classmin409"
## [1] "overlap5_classmin333"
## [1] "overlap5_classmin269"
## [1] "overlap5_classmin214"
## [1] "overlap5_classmin167"
## [1] "overlap5_classmin125"
## [1] "overlap5_classmin88"
## [1] "overlap5_classmin56"
## [1] "overlap5_classmin26"
## [1] "overlap5_classmin5"
```

Os mesmos podem ser carregados de forma semalhante as bases reais:

```
  DataSetName <- "overlap2_classmin88"
  ds <- GetDataSetMarcio(DataSetName)
```

```
## Parsed with column specification:
## cols(
##   X1 = col_double(),
##   X2 = col_double(),
##   Y = col_integer()
## )
```

```
  X <- ds$X
  Y <- ds$Y

  DataSetDesc <- GetDatasetdesc(DataSetName)

  head(X, 10)
```

```
##                X1         X2
##  [1,]  2.90716256  1.3029770
##  [2,]  2.14487689  2.8895576
##  [3,]  0.23546926  3.0854938
##  [4,] -0.45743250  1.1980162
##  [5,] -0.09346888  1.1437736
##  [6,]  1.29524122  1.0016366
##  [7,]  1.00688594  1.5891096
##  [8,]  2.15741089 -0.7729072
##  [9,]  3.13463789  0.5968447
## [10,]  1.23784461  0.9794085
```

```
    head(Y, 10)
```

```
##  [1] 1 1 1 1 1 1 1 1 1 1
## Levels: 0 1
```