Package ‘mfds’

September 8, 2017

Type Package

Title Multivariate Functional Data Sets

Version 0.1.0

Author Tomasz Gorecki, Lukasz Smaga

Maintainer Tomasz Gorecki <tomasz.gorecki@amu.edu.pl>

Description The package includes fifteen labeled multivariate functional data sets. The data sets were created from multivariate time series data available in the literature by extending all variables to the same length. They originate from different domains, including handwriting recognition, medicine, robotics, etc. The data sets can be used for illustrating and evaluating practical efficiency of classification and statistical inference methods.

License LGPL-2 | LGPL-3 | GPL-2 | GPL-3

Depends R (>= 2.10)

Encoding UTF-8

LazyData true

RoxygenNote 6.0.1

R topics documented:

`ArabicDigits` *Mel frequency cepstrum coefficients corresponding to spoken Arabic digits*

Description

This dataset contains time series of 13 Mel Frequency Cepstrum Coefficients (MFCCs) (Sampling rate: 11025 Hz, 16 bits, Window applied: hamming, Filter pre-emphasized: 1-0.97Z^(-1)) corresponding to spoken Arabic digits. Includes data from 44 males and 44 females native Arabic speakers between the ages 18 and 40 to represent ten spoken Arabic digit. The number of the time series is 8800 (10 digits x 10 repetitions x 88 speakers) in total. First half of each class contains male speakers whereas the second half contains female speakers.

Usage

```
ArabicDigits
data(ArabicDigits)
```

Format

A list of length 14 with the following components:

`MFCC1`, ..., `MFCC13` matrices with 93 rows and 8800 columns representing 13 Mel Frequency Cepstrum Coefficients.

`class` a factor of length 8800 with levels `"0"`, `"1"`, ..., `"9"` (ten spoken Arabic digit).

Note

The samples in this data set were originally of different lengths. They were extended to the length of the longest sample in the data set (Gorecki and Luczak, 2015).

Source

Lichman M (2013). UCI Machine Learning Repository [`http://archive.ics.uci.edu/ml`]. Irvine, CA: University of California, School of Information and Computer Science.

References

Gorecki T, Luczak M (2015). Multivariate Time Series Classification with Parametric Derivative Dynamic Time Warping. Expert Systems with Applications 42, 2305-2312.

Hammami N, Bedda M (2010). Improved Tree Model for Arabic Speech Recognition. Proc. IEEE ICCSIT10 Conference.

Examples

```
data(ArabicDigits)
str(ArabicDigits)
```

`AustralianLanguage` *Australian Sign Language Multivariate Functional Data*

Description

This data consists of sample of Auslan (Australian Sign Language) signs. 27 examples of each of 95 Auslan signs were captured (a total of 2565 signs were collected) from a native signer using high-quality position trackers and instrumented gloves and were collected over a period of nine weeks. The average length of each sign was approximately 57 frames.

The following data were recorded for each hand:

x position expressed relative to a zero point set slightly below the chin (in meters)

y position expressed relative to a zero point set slightly below the chin (in meters)

z position expressed relative to a zero point set slightly below the chin (in meters)

roll expressed as a value between -0.5 and 0.5 with 0 being palm down. Positive means the palm is rolled clockwise from the perspective of the signer. To get degrees, multiply by 180.

pitch expressed as a value between -0.5 and 0.5 with 0 being palm flat (horizontal). Positive means the palm is pointing up. To get degrees, multiply by 180.

yaw expressed a value between -1.0 and 1.0 with 0 being palm straight ahead from the perspective of the signer. Positive means clockwise from the perspective above the signer. To get degrees, multiply by 180.

Thumb bend measure between 0 and 1. 0 means totally flat, 1 means totally bent.

Forefinger bend measure between 0 and 1. 0 means totally flat, 1 means totally bent.

Middle finger bend measure between 0 and 1. 0 means totally flat, 1 means totally bent.

Ring finger bend measure between 0 and 1. 0 means totally flat, 1 means totally bent.

Little finger bend measure between 0 and 1. 0 means totally flat, 1 means totally bent.

Usage

```
AustralianLanguage
data(AustralianLanguage)
```

Format

A list of length 23 with the following components:

`LH_x`, ..., `RH_little_finger` matrices with 136 rows and 2565 columns.

`class` a factor of length 2565 with 95 levels (Auslan signs).

Note

The samples in this data set were originally of different lengths. They were extended to the length of the longest sample in the data set (Gorecki and Luczak, 2015).

Source

Lichman M (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

References

Gorecki T, Luczak M (2015). Multivariate Time Series Classification with Parametric Derivative Dynamic Time Warping. Expert Systems with Applications 42, 2305-2312.

Kadous MW (2002). Temporal Classification: Extending the Classification Paradigm to Multivariate Time Series. PhD Thesis, School of Computer Science and Engineering, University of New South Wales.

Examples

```
data(AustralianLanguage)
str(AustralianLanguage)
```

`CharacterTrajectories` *Character Trajectories Multivariate Functional Data*

Description

The data consists of 2858 character samples. The data was captured using a WACOM tablet. 3 Dimensions were kept: x, y, and pen tip force. The data has been numerically differentiated and Gaussian smoothed, with a sigma value of 2. Data was captured at 200Hz. The data was normalised. All samples are from the same writer. Only characters with a single pen-down segment were considered. Character segmentation was performed using a pen tip force cut-off point. The characters have also been shifted so that their velocity profiles best match the mean of the set. Each character sample is a 3-dimensional pen tip velocity trajectory.

Usage

```
CharacterTrajectories
data(CharacterTrajectories)
```

Format

A list of length 4 with the following components:

`x`, `y`, `pen tip force` matrices with 205 rows and 2858 columns representing the dimensions.

`class` a factor of length 2858 with 20 levels `"a"`, `"b"`, ..., `"z"`.

Note

The samples in this data set were originally of different lengths. They were extended to the length of the longest sample in the data set (Gorecki and Luczak, 2015).

Source

Lichman M (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

References

Gorecki T, Luczak M (2015). Multivariate Time Series Classification with Parametric Derivative Dynamic Time Warping. Expert Systems with Applications 42, 2305-2312.

Williams BH, Toussaint M, Storkey AJ (2006). Extracting Motion Primitives from Natural Handwriting Data. In ICANN 2, 634-643.

Examples

```
data(CharacterTrajectories)
str(CharacterTrajectories)
```

`ECG` *Measurements recorded by electrode during heartbeat*

Description

The ECG database comprises a collection of time series data sets where each file contains the sequence of measurements recorded by one electrode during one heartbeat. Each heartbeat has an assigned classification of normal or abnormal. All abnormal heartbeats are representative of a cardiac pathology known as supraventricular premature beat (SVPB). The ECG dataset contains 200 samples, among which 133 samples are normal and 67 samples are abnormal.

Usage

```
ECG
data(ECG)
```

Format

A list of length 3 with the following components:

`Lead_0`, `Lead_1` matrices with 152 rows and 200 columns.

`class` a factor of length 200 with 2 levels `"abnormal"` and `"normal"`.

Note

The samples in this data set were originally of different lengths. They were extended to the length of the longest sample in the data set (Gorecki and Luczak, 2015).

Source

Available from: `http://www.cs.cmu.edu/~bobski`

References

Gorecki T, Luczak M (2015). Multivariate Time Series Classification with Parametric Derivative Dynamic Time Warping. Expert Systems with Applications 42, 2305-2312.

Olszewski RT (2001). Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA.

Examples

```
data(ECG)
str(ECG)
```

Graz *Graz Multivariate Functional Data*

Description

This data set consists of EEG data from 9 subjects of a study published in Leeb et al. (2007). The subjects were right-handed, had normal or corrected-to-normal vision and were paid for participating in the experiments. All volunteers were sitting in an armchair, watching a flat screen monitor placed approximately 1 m away at eye level. For each subject 5 sessions are provided, whereby the first two sessions contain training data without feedback (screening), and the last three sessions were recorded with feedback. Three bipolar recordings (C3, Cz, and C4) were recorded with a sampling frequency of 250 Hz. The placement of the three bipolar recordings were slightly different for each subject. The cue-based screening paradigm consisted of two classes, namely the motor imagery (MI) of left hand (class 1) and right hand (class 2). Each subject participated in two screening sessions without feedback recorded on two different days within two weeks. Each session consisted of six runs with ten trials each and two classes of imagery. This resulted in 20 trials per run and 120 trials per session. Data of 120 repetitions of each MI class were available for each person in total.

Usage

```
Graz
data(Graz)
```

Format

A list of length 4 with the following components:

`C3`, `Cz`, `C4` matrices with 1152 rows and 140 columns representing bipolar recordings.

`class` a factor of length 140 with 2 levels `"LH"` and `"RH"` (motor imagery of left hand and right hand).

Source

Available from: http://www.bbci.de/competition/iv/

References

Leeb R, Lee F, Keinrath C, Scherer R, Bischof H, Pfurtscheller G (2007). Brain-Computer Communication: Motivation, Aim, and Impact of Exploring a Virtual Apartment. IEEE Transactions on Neural Systems and Rehabilitation Engineering 15, 473-482.

Examples

```
data(Graz)
str(Graz)
```

`JapaneseVowels` *Japanese Vowels Multivariate Functional Data*

Description

Nine male speakers uttered two Japanese vowels /ae/ successively. For each utterance, it was applied 12-degree linear prediction analysis (Sampling rate: 10kHz, Frame length: 25.6ms, Shift length: 6.4ms) to obtain a discrete-time series with 12 LPC cepstrum coefficients. This means that one utterance by a speaker forms a time series whose length is in the range 7-29 and each point of a time series is of 12 features (12 coefficients). The number of the time series is 640 in total.

Usage

```
JapaneseVowels
data(JapaneseVowels)
```

Format

A list of length 13 with the following components:

`LPC1`**, ...,** `LPC12` matrices with 29 rows and 640 columns representing 12 LPC cepstrum coefficients.

`class` a factor of length 640 with 9 levels `"Speaker 1"`, ..., `"Speaker 9"`.

Note

The samples in this data set were originally of different lengths. They were extended to the length of the longest sample in the data set (Gorecki and Luczak, 2015).

Source

Lichman M (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

References

Gorecki T, Luczak M (2015). Multivariate Time Series Classification with Parametric Derivative Dynamic Time Warping. Expert Systems with Applications 42, 2305-2312.

Kudo M, Toyama J, Shimbo M (1999). Multidimensional Curve Classification Using Passing-Through Regions. Pattern Recognition Letters 20, 1103-1111.

Examples

```
data(JapaneseVowels)
str(JapaneseVowels)
```

`Libras` *Official brazilian signal language*

Description

The data set contains 360 instances (15 classes of 24 instances each). Each class references to a hand movement type in LIBRAS (Portuguese name 'Lingua BRAsileira de Sinais', official brazilian signal language). The hand movement is represented as a bidimensional curve performed by the hand in a period of time. The curves were obtained from videos of hand movements. In the video pre-processing, a time normalization is carried out selecting 45 frames from each video, in according to a uniform distribution. In each frame, the centroid pixels of the hand are found, which compose the discrete version of the curve with 45 points. All curves are normalized in the unitary space. Each instance represents 45 points on a bidimensional space, which can be plotted in an ordered way to draw the path of the movement.

Usage

```
Libras
data(Libras)
```

Format

A list of length 3 with the following components:

`Coordinate abcissa`, `Coordinate ordinate` matrices with 45 rows and 360 columns.

`class` a factor of length 360 with 15 levels `"anti-clockwise arc"`, `"circle"`, ..., `"vertical zigzag"` (hand movement types in LIBRAS).

Source

Lichman M (2013). UCI Machine Learning Repository [`http://archive.ics.uci.edu/ml`]. Irvine, CA: University of California, School of Information and Computer Science.

References

Dias DB, Madeo RCB, Rocha T, Biscaro HH, Peres SM (2009). Hand Movement Recognition for Brazilian Sign Language: A Study Using Distance-Based Neural Networks. Proceedings of 2009 International Joint Conference on Neural Networks. Eau Claire, WI, USA: Documation LLC, 697-704.

Examples

```
data(Libras)
str(Libras)
```

PenDigits *Pen Digits Multivariate Functional Data*

Description

The data set contains 250 samples from 44 writers. It was used a WACOM PL-100V pressure sensitive tablet with an integrated LCD display and a cordless stylus. The tablet sends x and y tablet coordinates and pressure level values of the pen at fixed time intervals (sampling rate) of 100 milliseconds. Writers are asked to write 250 digits in random order inside boxes of 500 by 500 tablet pixel resolution. The first ten digits are ignored because most writers are not familiar with this type of input devices, but subjects are not aware of this. The stylus pressure level values were ignored. The dataset contains 10992 samples, divided into 10 groups (digits).

Usage

```
PenDigits
data(PenDigits)
```

Format

A list of length 3 with the following components:

`x`, `y` matrices with 8 rows and 10992 columns representing tablet coordinates.

`class` a factor of length 10992 with 10 levels `"0"`, `"1"`, ..., `"9"` (digits).

Source

Lichman M (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

References

Alimoglu F (1996). Combining Multiple Classifiers for Pen-Based Handwritten Digit Recognition. MSc Thesis, Institute of Graduate Studies in Science and Engineering, Bogazici University.

Examples

```
data(PenDigits)
str(PenDigits)
```

RobotFailure *Force and torque measurements on a robot after failure detection*

Description

This dataset contains force and torque measurements on a robot after failure detection. Each failure is characterized by 15 force/torque samples collected at regular time intervals starting immediately after failure detection. The donation includes 5 data sets, each of them defining a different learning problem:

LP1: failures in approach to grasp position,

LP2: failures in transfer of a part,

LP3: position of part after a transfer failure,

LP4: failures in approach to ungrasp position,

LP5: failures in motion with part.

The total observation window for each failure instance was of 315ms.

Usage

```
RobotFailure
data(RobotFailure)
```

Format

A list of length 5 with the following components:

`LP1` a list of length 7 with the following components:

- `X1`**, ...,** `X6` matrices with 15 rows and 88 columns.
- **class** a factor of length 88 with 4 levels `"collision"`, `"front collision"`, `"normal"`, `"obstruction"`.

`LP2` a list of length 7 with the following components:

- `X1`**, ...,** `X6` matrices with 15 rows and 47 columns.
- **class** a factor of length 47 with 5 levels `"back collision"`, `"collision to the left"`, `"collision to the right"`, `"front collision"`, `"ok"`.

`LP3` a list of length 7 with the following components:

- `X1`**, ...,** `X6` matrices with 15 rows and 47 columns.
- **class** a factor of length 47 with 4 levels `"lost"`, `"moved"`, `"ok"`, `"slightly moved"`.

`LP4` a list of length 7 with the following components:

- `X1`**, ...,** `X6` matrices with 15 rows and 117 columns.
- **class** a factor of length 117 with 3 levels `"collision"`, `"normal"`, `"obstruction"`.

`LP5` a list of length 7 with the following components:

- `X1`**, ...,** `X6` matrices with 15 rows and 164 columns.
- **class** a factor of length 164 with 5 levels `"bottom collision"`, `"bottom obstruction"`, `"collision in part"`, `"collision in tool"`, `"normal"`.

Source

Lichman M (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

References

Camarinha-Matos LM, Seabra Lopes L, Barata J (1996). Integration and Learning in Supervision of Flexible Assembly Systems. IEEE Transactions on Robotics and Automation 12, 202-219.

Examples

```
data(RobotFailure)
str(RobotFailure)
```

`uWaveGestureLibrary` *Acceleration data of gesture samples*

Description

Acceleration (3D) data of 4478 gesture samples are collected from eight users over an elongated period of time for a gesture vocabulary with eight gesture patterns identified by a Nokia research. For a participant, gestures are collected from seven days within a period of about three weeks. On each day, the participant holds the Wii remote in hand and repeats each of the eight gestures in the Nokia vocabulary ten times.

Usage

```
uWaveGestureLibrary
data(uWaveGestureLibrary)
```

Format

A list of length 4 with the following components:

`X`, `Y`, `Z` matrices with 315 rows and 4478 columns.

`class` a factor of length 4478 with 8 levels `"Gesture 1"`, ..., `"Gesture 8"` (users).

Source

Chen Y, Keogh E, Hu B, Begum N, Bagnall A, Mueen A, Batista G (2015). The UCR Time Series Classification Archive. URL www.cs.ucr.edu/~eamonn/time_series_data/

References

Liu J, Wang Z, Zhong L, Wickramasuriya J, Vasudevan V (2009). uWave: Accelerometer-Based Personalized Gesture Recognition and Its Applications. In Proc. IEEE Int. Conf. Pervasive Computing and Communication (PerCom).

Examples

```
data(uWaveGestureLibrary)
str(uWaveGestureLibrary)
```

`Wafer` *Measurements recorded by vacuum-chamber sensor during the etch process applied to silicon wafer during the manufacture of semiconductor microelectronics*

Description

The wafer database comprises a collection of time-series data sets where each file contains the sequence of measurements recorded by one vacuum-chamber sensor during the etch process applied to one silicon wafer during the manufacture of semiconductor microelectronics. The data sets were analyzed by appropriate domain experts, and a label of normal or abnormal was assigned to each data set. Of the 1194 data sets in the wafer database, 1067 data sets were identified as normal and 127 data sets were identified as abnormal. Six parameters have been identified as being critical for monitoring purposes: radio frequency forward power, radio frequency reflected power, chamber pressure, 405 nanometer (nm) emission, 520 nanometer (nm) emission, and direct current bias. The first two parameters are measures of electrical power applied to the plasma, the third parameter measures the pressure within the etch chamber, the fourth and fifth parameters measure the intensity of two different wavelengths (i.e., colors) of light emitted by the plasma, and the sixth parameter measures the direct current electrical potential difference within the tool.

Usage

```
Wafer
data(Wafer)
```

Format

A list of length 7 with the following components:

`RF_F_Power`, `RF_R_Power`, `Pressure`, `Emission_405`, `Emission_520`, `Bias` matrices with 198 rows and 1194 columns.

`class` a factor of length 1194 with 2 levels `"abnormal"`, `"normal"`.

Note

The samples in this data set were originally of different lengths. They were extended to the length of the longest sample in the data set (Gorecki and Luczak, 2015).

Source

Available from: http://www.cs.cmu.edu/~bobski

References

Gorecki T, Luczak M (2015). Multivariate Time Series Classification with Parametric Derivative Dynamic Time Warping. Expert Systems with Applications 42, 2305-2312.

Olszewski RT (2001). Generalized Feature Extraction for Structural Pattern Recognition in Time-Series Data. Ph.D. Thesis, Carnegie Mellon University, Pittsburgh, PA.

Examples

```
data(Wafer)
str(Wafer)
```

Index