



Beth Israel Deaconess
Medical Center



A teaching hospital of
Harvard Medical School

Workshop

Tools in Machine Learning

Albert C. Yang, M.D., Ph.D.

Beth Israel Deaconess Medical Center
Harvard Medical School, Boston, USA

Apr 9, 2018

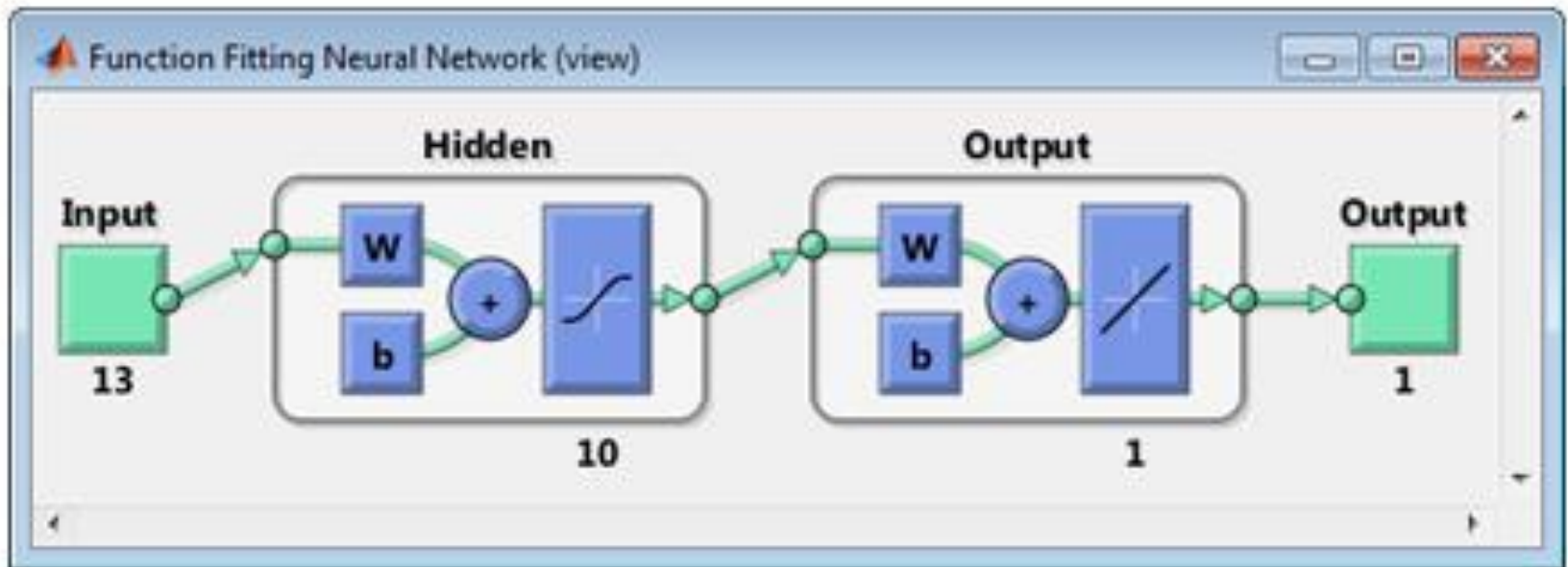
cyang1@bidmc.harvard.edu

Tools in Machine Learning

- Matlab
- Python
- R

Matlab

- Neural Network Toolbox



Python



[Don't Miss AnacondaCon Apr 8-11 Austin TX!](#)

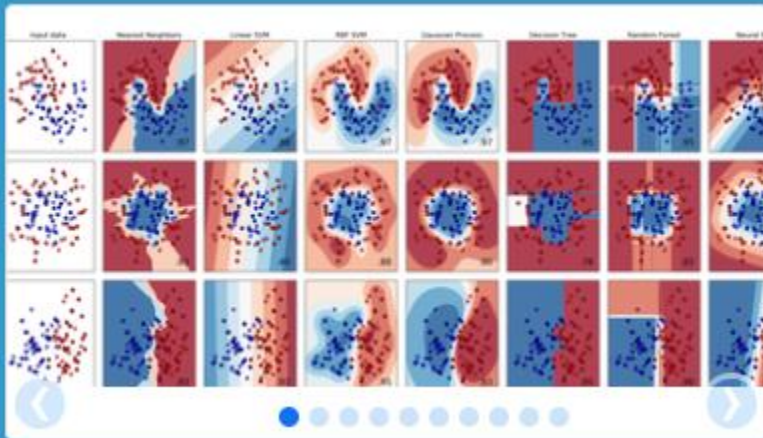
Download Anaconda Distribution

Version 5.1 | Release Date: February 15, 2018

Download For:   

Python

- Scikit-learn



scikit-learn

Machine Learning in Python

- Simple and efficient tools for data mining and data analysis
- Accessible to everybody, and reusable in various contexts
- Built on NumPy, SciPy, and matplotlib
- Open source, commercially usable - BSD license

Deep Learning in Python

Keras: The Python Deep Learning library



Keras

R



[\[Home\]](#)

Download

[CRAN](#)

R Project

[About R](#)

[Logo](#)

[Contributors](#)

[What's New?](#)

[Reporting Bugs](#)

[Development Site](#)

[Conferences](#)

[Search](#)

R Foundation

[Foundation](#)

[Board](#)

[Members](#)

[Donors](#)

[Donate](#)

The R Project for Statistical Computing

Getting Started

R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).

If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked questions](#) before you send an email.

News


- [R version 3.5.0 \(Joy in Playing\) prerelease versions](#) will appear starting Friday 2018-03-23. Final release is scheduled for Monday 2018-04-23.
- [R version 3.4.4 \(Someone to Lean On\)](#) has been released on 2018-03-15.
- [useR! 2018](#) (July 10 - 13 in Brisbane) is open for registration at <https://user2018.r-project.org>
- [The R Journal Volume 9/2](#) is available.
- [R version 3.3.3 \(Another Canoe\)](#) has been released on Monday 2017-03-06.
- [useR! 2017](#) took place July 4 - 7 in Brussels <https://user2017.brussels>
- The [R Logo](#) is available for download in high-resolution PNG or SVG formats.

Deep Learning in R

H2O Deep Learning, @ArnoCandel 11

H₂O + R = Happy Data Scientist

Machine Learning on Big Data with R:
Data resides on the H₂O cluster!



```
1 if ("package:h2o" %in% search()) { detach("package:h2o", unload=TRUE) }
2 if ("h2o" %in% rownames(installed.packages())) { remove.packages("h2o") }
3 install.packages("h2o", repos=c("file:///Users/arno/h2o/target/R", getOption("repos")))
4 library("h2o")
5 h2o_server = h2o.init()
6 %>% h2o.deeplearning
7 iris_train = h2o.importFile(h2o_server, path = '/Users/arno/h2o/smalldata/iris/iris.csv', header = 1, sep = ',', key = 'iris')
8 iris_model = h2o.deeplearning(x = 1:4, y = 5, data = iris_train, activation = "Tanh", hidden = c(50, 50, 50), epochs = 500)
9 iris_model
10 rns_100_train = h2o.importFile(h2o_server, path = '/Users/arno/rs/rns_100_deep_train.csv', header = T, sep = ',', key = 'rns_100_train')
11 rns_100_model = h2o.deeplearning(x = 2:101, y = 1, data = rns_100_train, classification = 1, activation = "Tanh", hidden = c(50, 50, 50))
12 rns_100_model
13 prediction = h2o.predict(rns_100_model, newdata = rns_100_train)
14 pred = as.data.frame(prediction)
15 head(pred)
16 tail(pred)
17 per = h2o.performance(prediction[,1], rns_100_train[,1], measure = "F1")
18 per
19 %>% h2o.performance
20 per@modelbest_cutoff
21
22
```

Environment History

Object	Class
pred	2031 obs. of 3 variables
iris_model	Formal class H2OClient
h2o_server	Formal class H2OClient
iris_train	Formal class H2ODeepLearning
rns_100_model	Formal class H2ODeepLearning
rns_100_train	Formal class H2OClient

Files Plots Packages Help Viewer

R Search Results

Search Results 

The search string was "performance"

Help pages:

- [h2o.performance](#) Build a Confusion Matrix from H2O Classification Predictions

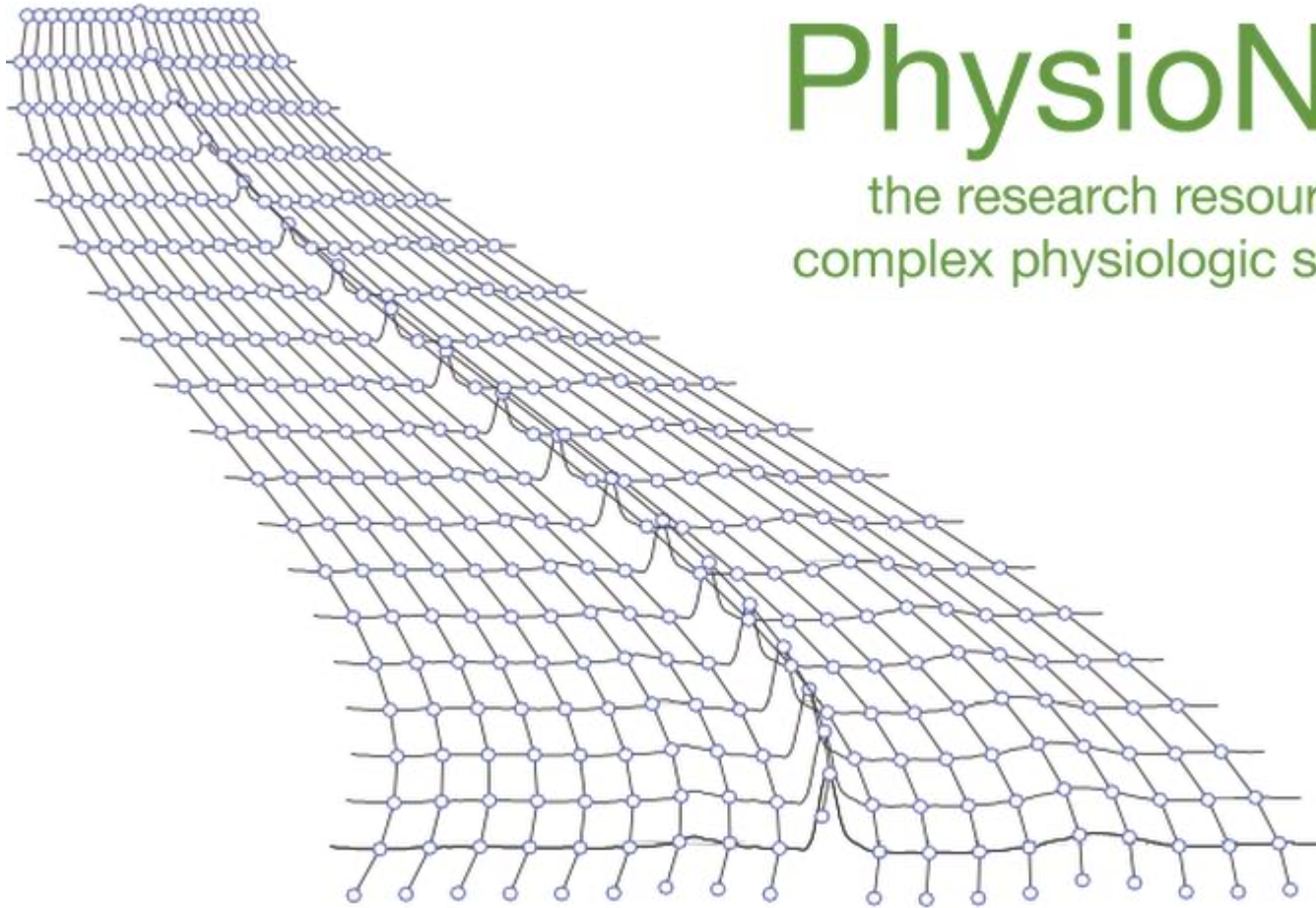
Resource of Sleep Data

- PhysioNet
- National Sleep Research Resource

PhysioNet

PhysioNet

the research resource for
complex physiologic signals



CAP Sleep Database

THE CAP SLEEP DATABASE

When referencing this material, please cite:

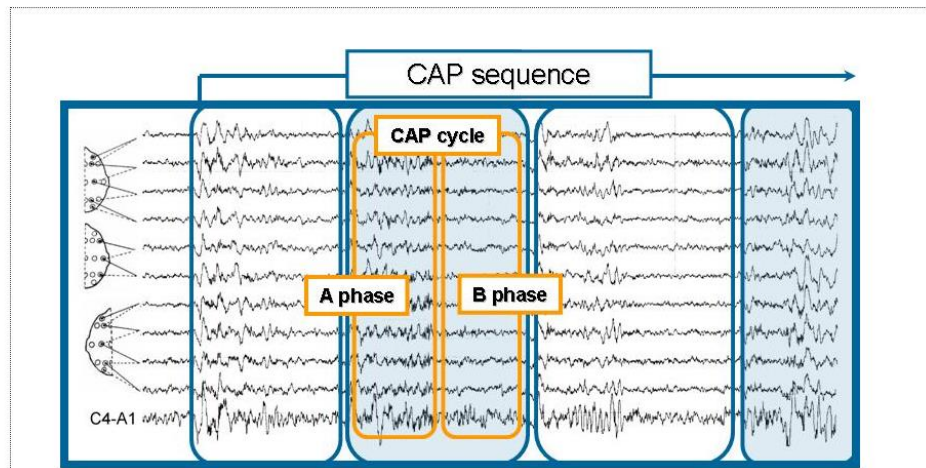
MG Terzano, L Parrino, A Sherieri, R Chervin, S Chokroverty, C Guilleminault, M Hirshkowitz, M Mahowald, H Moldofsky, A Rosa, R Thomas, A Walters. [Atlas, rules, and recording techniques for the scoring of cyclic alternating pattern \(CAP\) in human sleep](#). *Sleep Med* 2001 Nov; 2(6):537-553.

Please also include the standard citation for PhysioNet:

Goldberger AL, Amaral LAN, Glass L, Hausdorff JM, Ivanov PCh, Mark RG, Mietus JE, Moody GB, Peng C-K, Stanley HE. PhysioBank, PhysioToolkit, and PhysioNet: Components of a New Research Resource for Complex Physiologic Signals. *Circulation* 101(23):e215-e220 [Circulation Electronic Pages; <http://circ.ahajournals.org/cgi/content/full/101/23/e215>]; 2000 (June 13).

THE CYCLIC ALTERNATING PATTERN (CAP) OF EEG ACTIVITY DURING SLEEP

The Cyclic Alternating Pattern (CAP) is a periodic EEG activity occurring during NREM sleep. It is characterized by cyclic sequences of cerebral activation (phase A) followed by periods of deactivation (phase B) which separate two successive phase A periods with an interval <1 min. A phase A period and the following phase B period define a CAP cycle, and at least two CAP cycles are required to form a CAP sequence.



National Sleep Research Resource

National Sleep Research Resource

Free research data and tools.

[NSRR](#) [About](#) [Datasets](#) [Tools](#) [Forum](#) [Blog](#)

[Search](#) [Sign in](#)

What interests you?

POLYSOMNOGRAPHY



ACTIGRAPHY



DATASETS







SHARING



Not sure? View our [most popular](#) datasets.

National Sleep Research Resource

	All Ages ▾	All Data ▾	Publish Date ▾	
	SUBJECTS 5,804	AGES 40 - 89	TIMEFRAME 1995 - 2010	DATA Polysomnography
Sleep Heart Health Study 25,636 Files · 351 GB 1,991 Variables				Data Access Approved ✓
	SUBJECTS 1,243	AGES 5 - 9	TIMEFRAME 2007 - 2012	DATA Polysomnography
Childhood Adenotonsillectomy Trial 5,167 Files · 969 GB 3,139 Variables				Request Data Access ▶
	SUBJECTS 318	AGES 45 - 75	TIMEFRAME 2010 - 2012	DATA Polysomnography
Heart Biomarker Evaluation in Apnea Treatment 1,223 Files · 55.5 GB 891 Variables				Request Data Access ▶
	SUBJECTS 735	AGES 6 - 88	TIMEFRAME 2001 - 2006	DATA Polysomnography
Cleveland Family Study 2,245 Files · 150 GB 3,038 Variables				Data Access Approved ✓

National Sleep Research Resource



SUBJECTS

461

AGES

65 - 89

TIMEFRAME

2002 - 2003

DATA

Polysomnography

[Study of Osteoporotic Fractures](#) | [1,430 Files · 54 GB](#) | [1,282 Variables](#)

[Request Data Access](#) ▶



SUBJECTS

2,911

AGES

65 - 89

TIMEFRAME

2003 - 2012

DATA

Polysomnography

[MrOS Sleep Study](#) | [11,858 Files · 1.18 TB](#) | [766 Variables](#)

[Request Data Access](#) ▶

*Cleveland Children's
Sleep and Health Study*

SUBJECTS

517

AGES

16 - 19

TIMEFRAME

2006 - 2010

DATA

Polysomnography

[Cleveland Children's Sleep and Health Study](#) | [1,589 Files · 214 GB](#) | [346 Variables](#)

[Request Data Access](#) ▶



SUBJECTS

16,415

AGES

18 - 76

TIMEFRAME

2009 - 2013

DATA

Polysomnography
Actigraphy

[Hispanic Community Health Study / Study of Latinos](#) | [26,094 Files · 32 GB](#) | [875 Variables](#)

[Request Data Access](#) ▶