



Workshop

Tools in Machine Learning

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

Institutes of Brain Science/Digital Medicine Center
National Yang-Ming University

Mar 19, 2019

accyang@gmail.com

Tools in Machine Learning

- Matlab
- Python
- R

Matlab



Products

Solutions

Academia

Support

Community

Events

Statistics and Machine Learning Toolbox

Search MathWorks.c

Statistics and Machine Learning Toolbox

Analyze and model data using statistics and machine learning

▶ Watch video

⬇️ Download a free trial

Matlab



Products

Solutions

Academia

Support

Community

Events

Deep Learning Toolbox

MAJOR UPDATE

Search MathWorks

Deep Learning Toolbox

Create, analyze, and train deep learning networks

▶ Watch video

⬇️ Download a free trial



Python



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

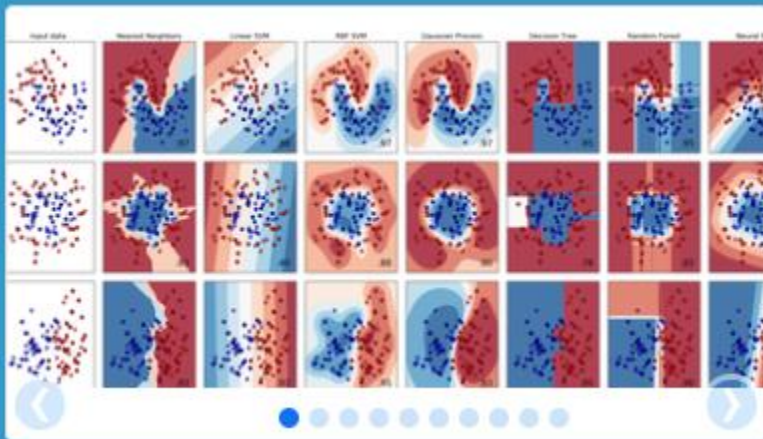
Download Anaconda Distribution

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
iris_train	Formal class H2OdeepLearningModel
h2o_server	Formal class H2OGBModel
rns_100_model	Formal class H2OClient
rns_100_train	Formal class H2OdeepLearningModel

Search Results

The search string was "h2o"

Help pages:

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

Importing Excel Data into Matlab

- `[num,txt,raw] = xlsread('Brain_Sex_Classification_YoungAdult_GSP_sMRI_600.xlsx');`

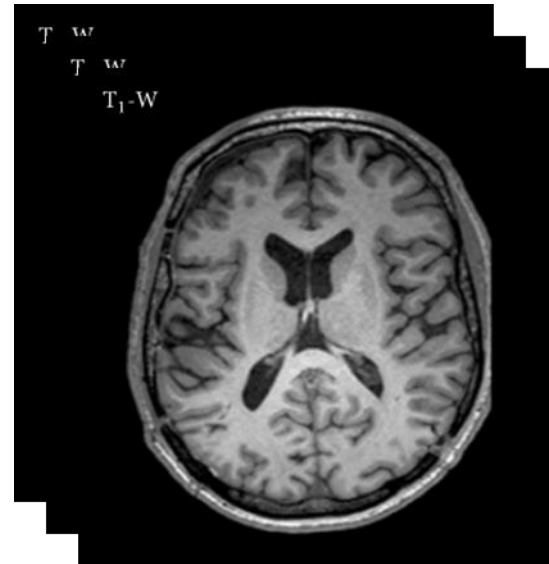
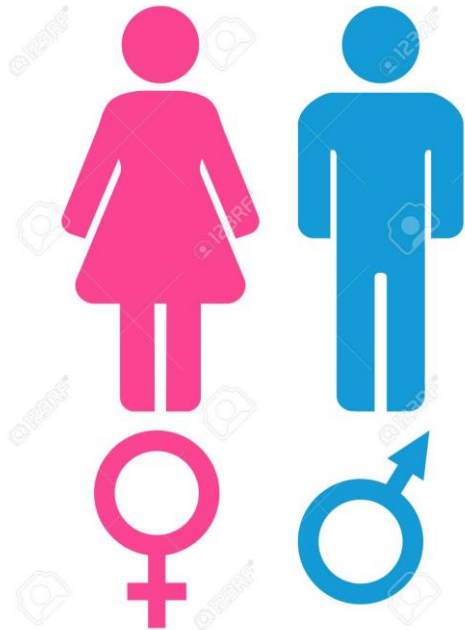


The screenshot shows the MATLAB Command Window and Workspace. The Command Window displays the command `[num,txt,raw] = xlsread('Brain_Sex_Classification_YoungAdult_GSP_sMRI_600.xlsx');` and the prompt `fx >> |`. The Workspace window shows the variables `num`, `raw`, and `txt` with their respective values: `num` is a `1200x45 double`, `raw` is a `1201x45 cell`, and `txt` is a `1201x45 cell`.

Name	Value
num	1200x45 double
raw	1201x45 cell
txt	1201x45 cell

Variables Organization

- `sex=num(:,3);`
- `brain=num(:,6:45);`



Look into Brain Metrics

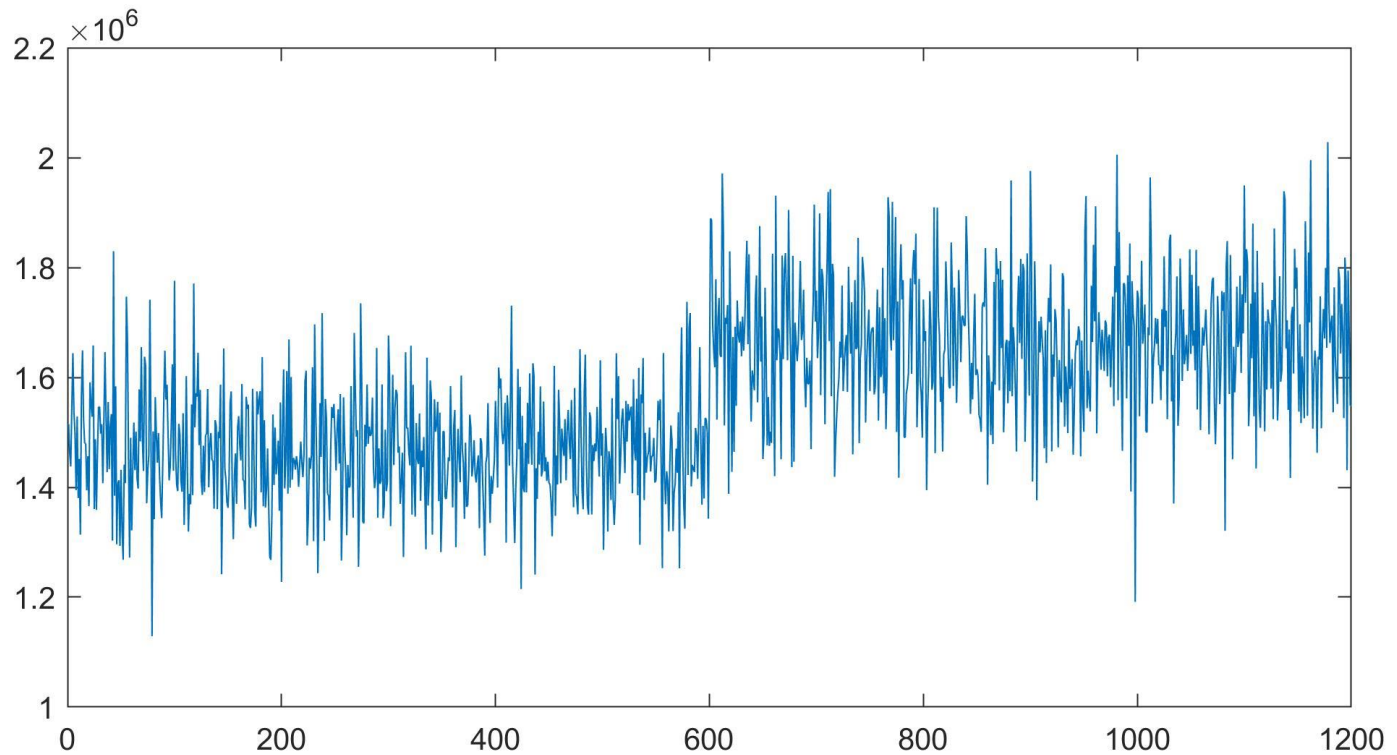
brain x

1200x40 double

	1	2	3	4	5	6	7	8	9
1	1.5150e+06	1189565	1177053	1071	661	601	556	913	2.36e+06
2	1.4632e+06	1144258	1130784	1054	574	489	425	945	2.52e+06
3	1.4379e+06	1126298	1117216	920	464	522	474	807	2.60e+06
4	1.5448e+06	1234031	1223971	960	593	774	715	1147	2.39e+06
5	1.6441e+06	1259277	1250056	962	610	626	633	866	2.35e+06
6	1.5365e+06	1248211	1237336	941	478	512	487	886	2.50e+06
7	1.4981e+06	1165208	1151957	1104	447	424	527	829	2.57e+06
8	1.3947e+06	1139605	1126518	918	550	509	495	925	2.46e+06
9	1.5288e+06	1175892	1164037	918	549	513	417	842	2.62e+06
10	1.3803e+06	1097100	1080649	868	395	375	366	705	2.48e+06
11	1.4517e+06	1169329	1156849	876	455	486	416	941	2.47e+06
12	1.3138e+06	1006265	993326	906	398	425	428	846	2.34e+06
13	1.5660e+06	1234050	1219934	915	448	531	560	875	2.47e+06
14	1.6491e+06	1269133	1258709	961	580	841	791	865	2.42e+06
15	1.5226e+06	1189312	1161984	893	323	313	312	668	2.51e+06
16	1.4803e+06	1161407	1145072	752	296	350	367	830	2.44e+06
17	1.4795e+06	1163231	1150676	1015	453	435	422	805	2.49e+06
18	1.3044e+06	1127640	1112452	1202	502	622	478	707	2.54e+06

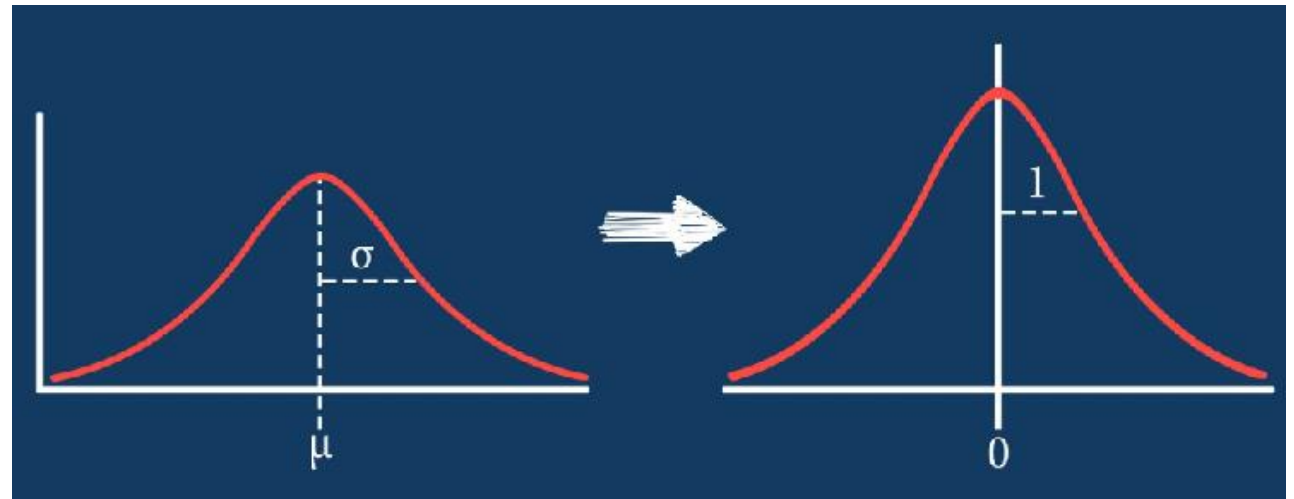
Look into Brain Metrics

- `plot(brain(:,1));`



Data Normalization

- `m = mean(brain);`
- `sd = std(brain);`
- `for i=1:40`
- `brain(:,i) = (brain(:,i)-m(i))/sd(i);`
- `end`



Normalized Data

brain x

1200x40 double

	1	2	3	4	5	6	7	8	9
1	-0.3498	-0.5498	-0.5490	1.1641	2.2142	1.0047	0.7419	0.3939	-1.860
2	-0.6954	-0.9133	-0.9260	1.0413	1.1861	-0.0760	-0.5637	0.6350	0.230
3	-0.8639	-1.0573	-1.0366	0.0734	-0.1138	0.2424	-0.0754	-0.4049	1.23
4	-0.1513	-0.1931	-0.1668	0.3623	1.4106	2.6740	2.3266	2.1572	-1.39
5	0.5110	0.0094	0.0458	0.3768	1.6115	1.2459	1.5093	0.0397	-2.02
6	-0.2062	-0.0794	-0.0579	0.2251	0.0516	0.1459	0.0542	0.1904	-0.08
7	-0.4625	-0.7452	-0.7535	1.4024	-0.3147	-0.7032	0.4529	-0.2391	0.88
8	-1.1519	-0.9506	-0.9608	0.0590	0.9025	0.1170	0.1339	0.4843	-0.56
9	-0.2576	-0.6595	-0.6551	0.0590	0.8907	0.1556	-0.6435	-0.1411	1.53
10	-1.2481	-1.2916	-1.3345	-0.3021	-0.9292	-1.1760	-1.1518	-1.1735	-0.28
11	-0.7718	-0.7121	-0.7137	-0.2443	-0.2202	-0.1049	-0.6534	0.6049	-0.37
12	-1.6911	-2.0203	-2.0460	-0.0277	-0.8938	-0.6935	-0.5338	-0.1110	-2.15
13	-0.0098	-0.1930	-0.1997	0.0373	-0.3029	0.3293	0.7818	0.1075	-0.42
14	0.5444	0.0885	0.1163	0.3696	1.2570	3.3205	3.0840	0.0322	-1.03
15	-0.2991	-0.5518	-0.6718	-0.1216	-1.7801	-1.7742	-1.6900	-1.4523	0.07
16	-0.5811	-0.7757	-0.8096	-1.1399	-2.0992	-1.4172	-1.1418	-0.2316	-0.83
17	-0.5863	-0.7611	-0.7640	0.7596	-0.2438	-0.5971	-0.5936	-0.4199	-0.15
18	1.1527	1.0466	1.0672	2.1102	1.4106	1.2072	0.0255	0.4802	0.42



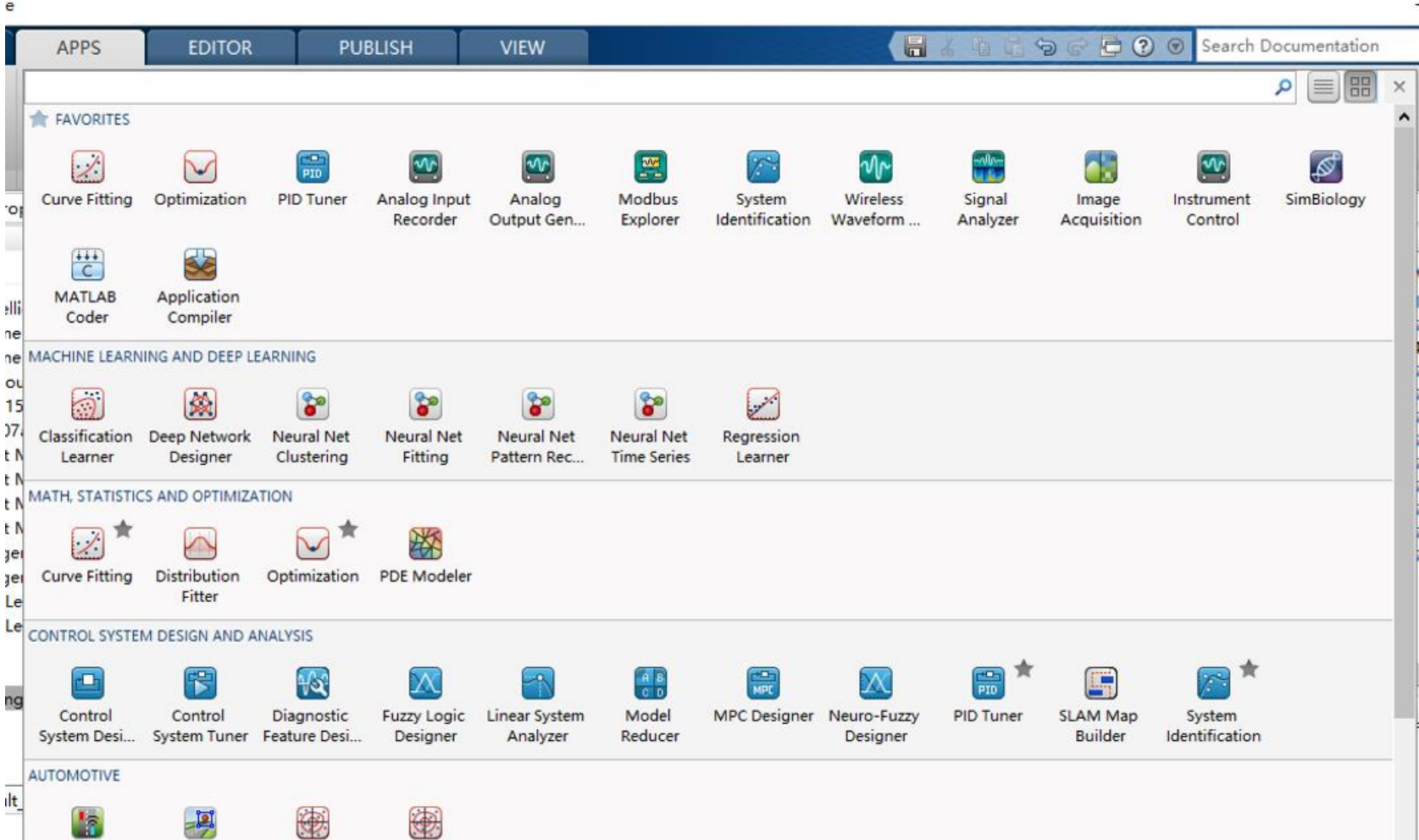
1200 Cases

40 Variables
40 Variables

Prepare for Machine Learning Data

- `mldata = [brain sex];`

Matlab Machine Learning APP



Linear Discrimination

