

H2O Machine Learning & Deep Learning London Workshop



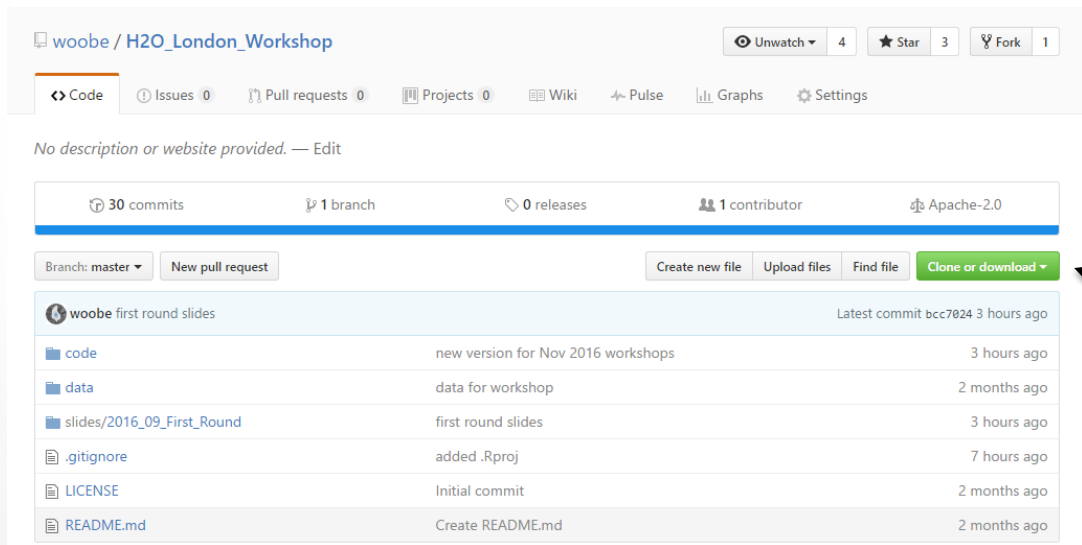
Jo-fai (Joe) Chow
Data Scientist
joe@h2o.ai
@matlabulous

Data Science for IoT Meetup
Barclays Eagle Venture Labs
21st & 24th November, 2016

Download Data & Code for Workshop

- Please go to

bit.ly/h2o_iot_workshop1



woobe / H2O_London_Workshop

Unwatch 4 Star 3 Fork 1

Code Issues 0 Pull requests 0 Projects 0 Wiki Pulse Graphs Settings

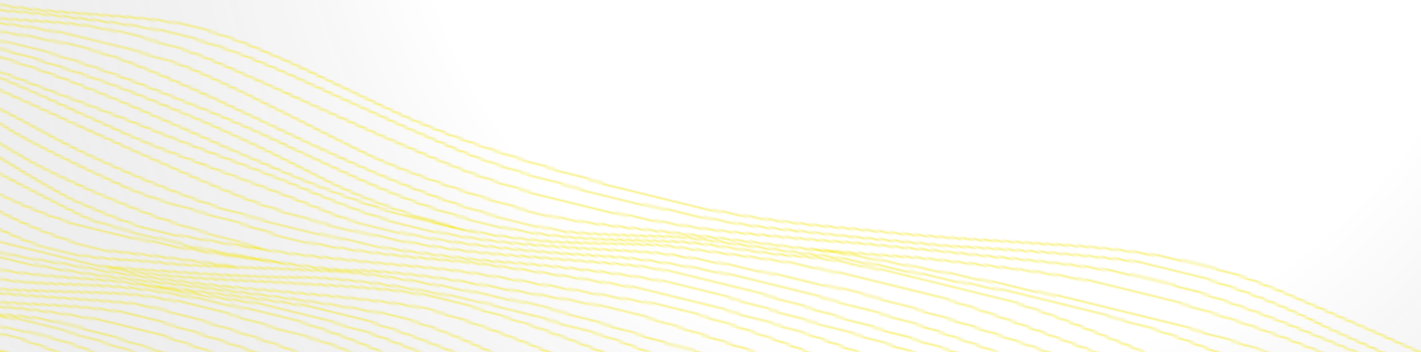
No description or website provided. — Edit

30 commits 1 branch 0 releases 1 contributor Apache-2.0

Branch: master New pull request Create new file Upload files Find file Clone or download

| woobe first round slides | Latest commit bcc7824 3 hours ago |
|----------------------------|--|
| code | new version for Nov 2016 workshops 3 hours ago |
| data | data for workshop 2 months ago |
| slides/2016_09_First_Round | first round slides 3 hours ago |
| .gitignore | added .Rproj 7 hours ago |
| LICENSE | Initial commit 2 months ago |
| README.md | Create README.md 2 months ago |

Example 2: Anomaly Detection



use_case_2_anomaly_detection.R

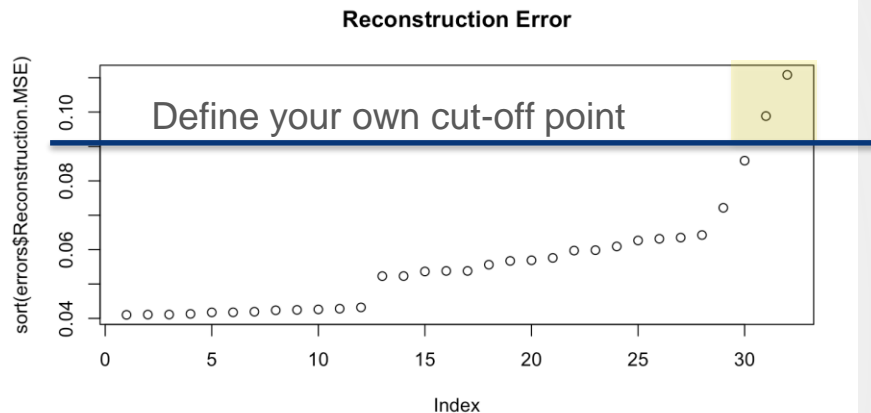
```
1 # -----  
2 # Step 8: Using Deep Learning for Anomaly Detection  
3 # -----  
4  
5 # Start and connect to a local H2O cluster  
6 library(h2o)  
7 h2o.init(nthreads = -1)  
8  
9 # Import data from a local CSV file  
10 mtcars <- read.csv("./data/auto_design.csv")  
11 mtcars$gear <- as.factor(mtcars$gear)  
12 mtcars$carb <- as.factor(mtcars$carb)  
13 mtcars$cyl <- as.factor(mtcars$cyl)  
14 mtcars$vs <- as.factor(mtcars$vs)  
15 mtcars$am <- as.factor(mtcars$am)  
16 mtcars$ID <- 1:nrow(mtcars)  
17  
18 # Print it out  
19 print(mtcars)  
20  
21 # Convert R data frame into H2O data frame  
22 h2o_mtcars <- as.h2o(mtcars)
```

```
> print(mtcars)
```

| | | X | mpg | cyl | disp | hp | drat | wt | qsec | vs | am | gear | carb | ID |
|----|--|---------------------|------|-----|-------|------|--------|---------|---------|----|----|------|------|----|
| 1 | | Mazda RX4 | 21.0 | 6 | 160.0 | 110 | 3.90 | 2.620 | 16.46 | 0 | 1 | 4 | 4 | 1 |
| 2 | | Mazda RX4 Wag | 21.0 | 6 | 160.0 | 110 | 3.90 | 2.875 | 17.02 | 0 | 1 | 4 | 4 | 2 |
| 3 | | Datsun 710 | 22.8 | 4 | 108.0 | 93 | 3.85 | 2.320 | 18.61 | 1 | 1 | 4 | 1 | 3 |
| 4 | | Hornet 4 Drive | 21.4 | 6 | 258.0 | 110 | 3.08 | 3.215 | 19.44 | 1 | 0 | 3 | 1 | 4 |
| 5 | | Hornet Sportabout | 18.7 | 8 | 360.0 | 175 | 3.15 | 3.440 | 17.02 | 0 | 0 | 3 | 2 | 5 |
| 6 | | Valiant | 18.1 | 6 | 225.0 | 105 | 2.76 | 3.460 | 20.22 | 1 | 0 | 3 | 1 | 6 |
| 7 | | Duster 360 | 14.3 | 8 | 360.0 | 245 | 3.21 | 3.570 | 15.84 | 0 | 0 | 3 | 4 | 7 |
| 8 | | Merc 240D | 24.4 | 4 | 146.7 | 62 | 3.69 | 3.190 | 20.00 | 1 | 0 | 4 | 2 | 8 |
| 9 | | Merc 230 | 22.8 | 4 | 140.8 | 95 | 3.92 | 3.150 | 22.90 | 1 | 0 | 4 | 2 | 9 |
| 10 | | Merc 280 | 19.2 | 6 | 167.6 | 123 | 3.92 | 3.440 | 18.30 | 1 | 0 | 4 | 4 | 10 |
| 11 | | Merc 280C | 17.8 | 6 | 167.6 | 210 | 800.00 | 900.000 | 1000.00 | 1 | 0 | 4 | 4 | 11 |
| 12 | | Merc 450SE | 16.4 | 8 | 275.8 | 180 | 3.07 | 4.070 | 17.40 | 0 | 0 | 3 | 3 | 12 |
| 13 | | Merc 450SL | 17.3 | 8 | 275.8 | 180 | 3.07 | 3.730 | 17.60 | 0 | 0 | 3 | 3 | 13 |
| 14 | | Merc 450SLC | 15.2 | 8 | 275.8 | 180 | 3.07 | 3.780 | 18.00 | 0 | 0 | 3 | 3 | 14 |
| 15 | | Cadillac Fleetwood | 10.4 | 8 | 472.0 | 205 | 2.93 | 5.250 | 17.98 | 0 | 0 | 3 | 4 | 15 |
| 16 | | Lincoln Continental | 10.4 | 8 | 460.0 | 215 | 3.00 | 5.424 | 17.82 | 0 | 0 | 3 | 4 | 16 |
| 17 | | Chrysler Imperial | 14.7 | 8 | 440.0 | 230 | 3.23 | 5.345 | 17.42 | 0 | 0 | 3 | 4 | 17 |
| 18 | | Fiat 128 | 32.4 | 4 | 780.0 | 2100 | 400.00 | 200.000 | 700.00 | 1 | 1 | 4 | 1 | 18 |
| 19 | | Honda Civic | 80.4 | 10 | 75.7 | 100 | 4.93 | 1.615 | 150.52 | 1 | 1 | 4 | 2 | 19 |
| 20 | | Toyota Corolla | 33.9 | 4 | 71.1 | 65 | 4.22 | 1.835 | 19.90 | 1 | 1 | 4 | 1 | 20 |
| 21 | | Toyota Corona | 21.5 | 4 | 120.1 | 97 | 3.70 | 2.465 | 20.01 | 1 | 0 | 3 | 1 | 21 |
| 22 | | Dodge Challenger | 15.5 | 8 | 318.0 | 150 | 2.76 | 3.520 | 16.87 | 0 | 0 | 3 | 2 | 22 |
| 23 | | AMC Javelin | 15.2 | 8 | 304.0 | 150 | 3.15 | 3.435 | 17.30 | 0 | 0 | 3 | 2 | 23 |
| 24 | | Camaro Z28 | 13.3 | 8 | 350.0 | 245 | 3.73 | 3.840 | 15.41 | 0 | 0 | 3 | 4 | 24 |
| 25 | | Pontiac Firebird | 19.2 | 8 | 400.0 | 175 | 3.08 | 3.845 | 17.05 | 0 | 0 | 3 | 2 | 25 |
| 26 | | Fiat X1-9 | 27.3 | 4 | 79.0 | 66 | 4.08 | 1.935 | 18.90 | 1 | 1 | 4 | 1 | 26 |
| 27 | | Porsche 914-2 | 26.0 | 4 | 120.3 | 91 | 4.43 | 2.140 | 16.70 | 0 | 1 | 5 | 2 | 27 |
| 28 | | Lotus Europa | 30.4 | 4 | 95.1 | 113 | 3.77 | 1.513 | 16.90 | 1 | 1 | 5 | 2 | 28 |
| 29 | | Ford Pantera L | 15.8 | 8 | 351.0 | 264 | 4.22 | 3.170 | 14.50 | 0 | 1 | 5 | 4 | 29 |
| 30 | | Ferrari Dino | 19.7 | 6 | 900.0 | 700 | 3.62 | 200.770 | 150.50 | 0 | 1 | 5 | 6 | 30 |
| 31 | | Maserati Bora | 15.0 | 8 | 301.0 | 335 | 3.54 | 3.570 | 14.60 | 0 | 1 | 5 | 8 | 31 |
| 32 | | Volvo 142E | 21.4 | 4 | 121.0 | 109 | 4.11 | 2.780 | 18.60 | 1 | 1 | 4 | 2 | 32 |

use_case_2_anomaly_detection.R

```
25 # -----
26 # Training an unsupervised deep neural network with autoencoder
27 # -----
28
29 # Use a bigger DNN
30 model <- h2o.deeplearning(x = 1:10,
31                           training_frame = h2o_mtcars,
32                           autoencoder = TRUE,
33                           activation = "RectifierWithDropout",
34                           hidden = c(50, 50, 50),
35                           epochs = 100)
36
37 # Calculate reconstruction errors (MSE)
38 errors <- h2o.anomaly(model, h2o_mtcars, per_feature = FALSE)
39 print(errors)
40 errors <- as.data.frame(errors)
41
42 # Plot
43 plot(sort(errors$Reconstruction.MSE), main = "Reconstruction Error")
44
45 # Outliers (define 0.09 as the cut-off point)
46 row_outliers <- which(errors > 0.09) # based on plot above
47 mtcars[row_outliers,]
```



```
> row_outliers <- which(errors > 0.09) # based on plot above
> mtcars[row_outliers,]
```

| | X | mpg | cyl | displacement | horsepower | weight | quarter mile time | vs | am | gear | carburetor | ID | | |
|----|------|------|------|--------------|------------|--------|-------------------|-----|------|------|------------|----|---|----|
| 11 | Merc | 280C | 17.8 | 6 | 167.6 | 210 | 800 | 900 | 1000 | 1 | 0 | 4 | 4 | 11 |
| 18 | Fiat | 128 | 32.4 | 4 | 780.0 | 2100 | 400 | 200 | 700 | 1 | 1 | 4 | 1 | 18 |

Identify the outliers

Thanks!

- Organisers & Contributors
 - Ajit Jaokar
 - Sibanjean Das
- Key Resources
 - docs.h2o.ai
 - github.com/h2oai/h2o-meetups
- Slides & Code
 - bit.ly/h2o_iot_workshop1
- Contact
 - joe@h2o.ai
 - @matlabulous
 - github.com/woobe