***Requests from FM from Report (10\_2\_2019) under General Issues:***

* *The Japanese Presentation:* ***FM asked to PC to send it to us****.*

(<http://yosinski.com/mlss12/media/slides/MLSS-2012-Sugiyama-Density-Ratio-Estimation-in-Machine-Learning.pdf>)

* *The MOA software:* ***FM asked to send us the link of the software.***

(<https://moa.cms.waikato.ac.nz/downloads/>)

* ***FM asked to PC to send it tonight or tomorrow morning considering that we have the NASA meeting on Friday (to have time to review it).***

**Problem statement:** Analyzing changes in evolving data from advanced habitation systems.

The goal of this task is to create a framework that monitors and provides change detection in a multimodal system. Our objective is two-fold (a) To provide the mining of patterns from older data as changes in data could reflect long term/ previous trends. (b) Mining for patterns over near data – as recent changes in data could indicate the recurrences of the previously known pattern or an upcoming event.

**Tasks:**

**T1:** High Level Cognitive Engine for Adaptable Diagnostic Design

**T2:** Semi-Autonomous Learning for Advanced Habitation Systems Monitoring

**T2.1:** Semi-Autonomous Embedded Collaborative Learning

**T2.2:** Retrospective CPD

**T4:** Distributed Human System Integration and Awareness Environment

**T4.1:** Change point detection and validation

**T4.2:** User Interface design

**Reported outcomes:**

**T1 – Appendix A:** Implementation of CNN using Tensor Flow that provides the high-level cognitive engine for adaptable diagnostic design, with the following specific aims:

1. Explore some of the TensorFlow capabilities[[1]](#footnote-1)
   1. Environments
   2. Exploratory Data Analysis using CIFAR images (known datasets)
   3. Create neural network architectures
   4. Create convolution neural network architectures
2. Create a CNN architecture to do a multi-feature extraction and high-level representation of the data[[2]](#footnote-2)

**T2.1 – (Details in Appendix B):** The use of sparklyr package in R to establish the spark connection, with the following specific aims:

1. Implemented Spark Session to integrate the csv file inputs with apache spark.
2. writing a SQL query to retrieve data.
3. Investigate the use of sparklyr package[[3]](#footnote-3) and tidyverse package[[4]](#footnote-4)

**T2.2 – (Details in Appendix C):** Based on the work of Liu et al (2012)[[5]](#footnote-5), we tested the with building block code for a simple sliding window on small multimodal data using the density-ratio estimation library[[6]](#footnote-6) , with the following specific aims:

1. Research and understand density-ratio estimation.
2. Implement density-ratio estimation on multi-modal data (data from MOA data generator).
3. Expand density-ratio estimation code to use uLSIF [[7]](#footnote-7)and *f-*divergence to detect changes in multi-modal data.

**T4.1 – (Details in Appendix D):** Simulating Concept Drift in MOA and analyzing the sequential analysis-based concept drift detection method, with the following specific aims:

1. Determining the issues while integrating R to MOA specifically with RMOA[[8]](#footnote-8).
2. Using MOA, comparing the evaluation measures of NaiveBayes and HoeffdingTree classifiers.
3. Analyzing the fundamentals and issues with the Cumulative Sum (CUSUM) algorithm.

**Appendix A**

**Results T1:**

First, some environments were tested, since some problems have been noticed in the past when working with Windows machines. Environments tested are: jupyter notebook on Macintosh machine, jupyter notebook on Windows machine, PyCharm on Macintosh machine, PyCharm on Windows machine, Google Colaboratory on Macintosh machine and Google Colaboratory on Windows machine.

The results were good on all cases in the Macintosh machine. In the Windows machine there were decreasing problems as the experiments migrated from the local machine to full cloud applications, which had no problems. The source of such problems still unknown.

Second, some tutorials and examples were followed.

Since this first part is not much of interest, just two examples will be talked about.

First example is of a neural network on a regression problem. It tries to predict fuel efficiency of early automobiles using the Auto MPG Dataset. The few unknown values were dropped out of the dataset for simplification purposes. The Origin feature was converted into a one-hot vector. 80% of the data was used as training and the rest, as testing.

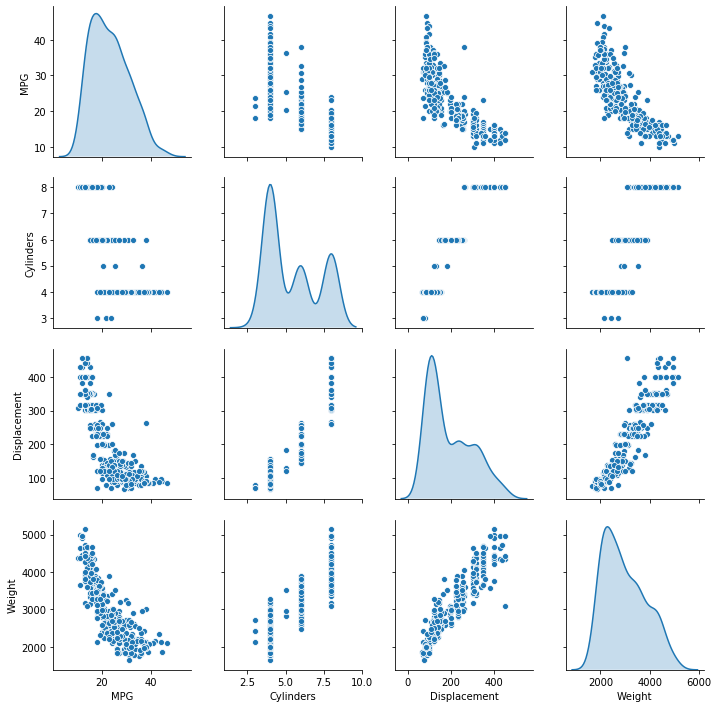


Fig 1 – Exploratory Data Analisys

On Fig 1 it is shown a feature in term of another, of all the features that should have some form of correlation. We can see that there is a correlation between the features, as expected.

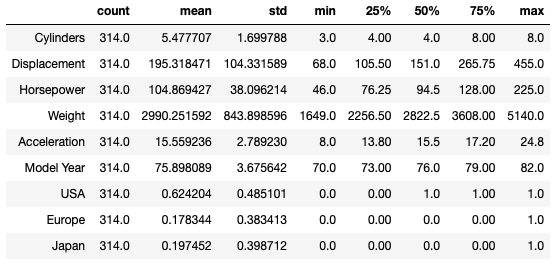


Fig 2 – Probability distribution of the features

Looking at the table, it can be observed that the data is varying from a feature to another. To obtain better results, normalization of the data was done. The normalization was done in this way: (x – mean(x))/std(x), were x is a feature.

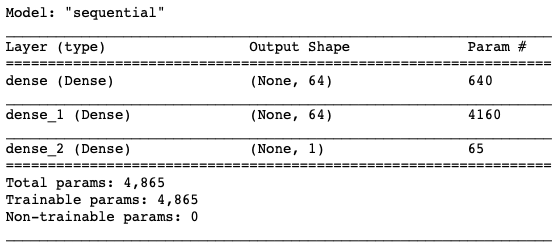


Fig 3 – Neural network architecture

Fig 3 shows how the neural network is setup. The input feeds to a fully connected layer of 64 neurons, which feeds to another fully connected layer of 64 neurons – both having a rectified linear unit as an activation function – and it all condenses as one output layer with a single, continuous value.

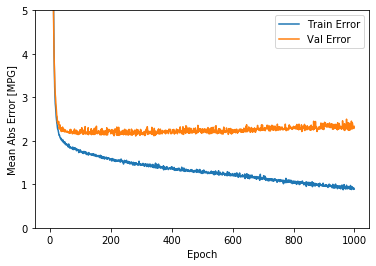


Fig 4 – Mean Absolute Error for the training data and for the validation data (1000 epochs)

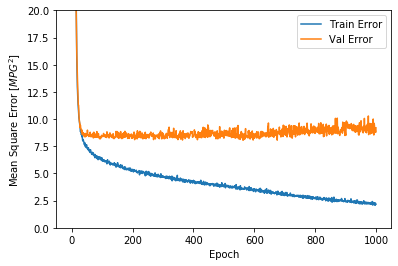


Fig 5 – Mean Square Error for the training data and for the validation data (1000 epochs)

As validation data shows, the model did not get much better after 100 epochs. It can be said that there was even some level of degradation of the model.

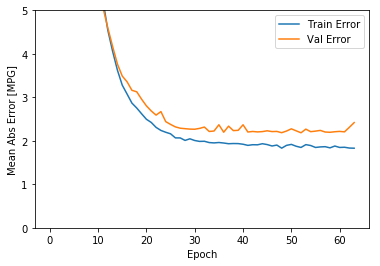


Fig 6 – Mean Absolute Error for the training data and for the validation data (1000 epochs, early\_stop activated)

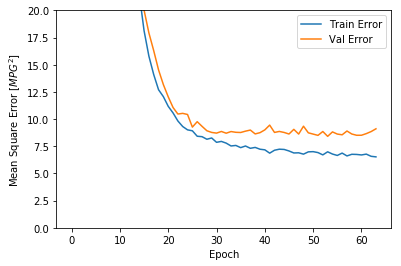


Fig 7 – Mean Square Error for the training data and for the validation data (1000 epochs, early\_stop activated)

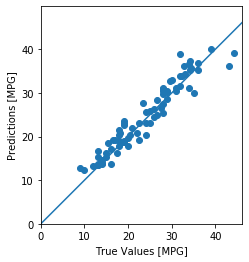


Fig 8 – Prediction of the regression model

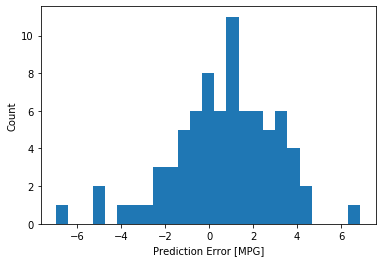


Fig 9 – Distribution of prediction error

As previously observed, the model does a good job until somewhere near the 60th epoch, when it starts to degrade, and its training is automatically stopped.

The prediction and the prediction error plots are made with the model being stopped before the 70th epoch.

It does a good job, as expected from the training and test error values and as shown in Fig 8. The distribution of prediction error is close to a Gaussian. A bigger dataset should confirm that.

Next, a convolutional neural network to classify CIFAR images. This dataset has 60,000 color images in 10 distinct classes without any ovelaping, with 6,000 images each class. It is divided into 50,000 training images and 10,000 testing images.

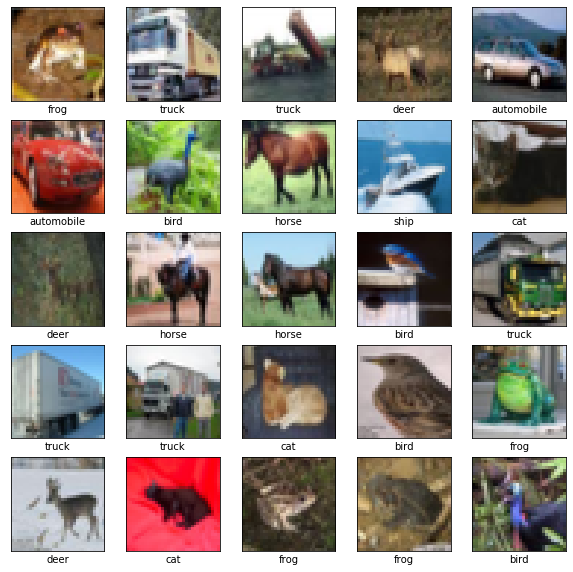


Fig 10 – Some samples from the training set, along with its respective label

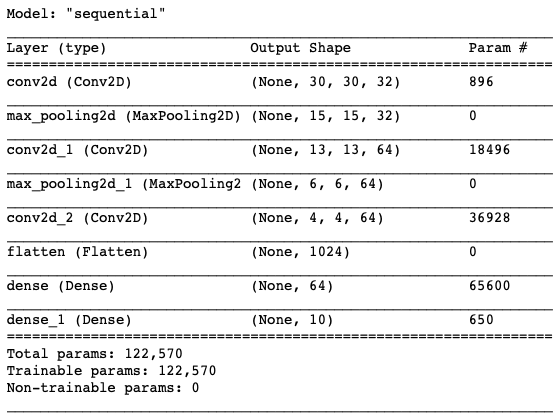


Fig 11 – The convolution neural network architecture

The CNN architecture consists of a convolution layer of 32 neurons with a rectified linear unit (ReLU) activation function, that feeds into a pooling layer of 2 by 2 (halves both the number of columns and the number of rows), repeats the same process again but the convolution having 64 neurons this time. After the second pooling, another convolution layer of 64 neurons with a ReLU activation function is performed, the results are flattened and fed to a last neural network with 64 neurons with a ReLU activation function, before being fed to an output layer of 10 neurons with a softmax activation function.

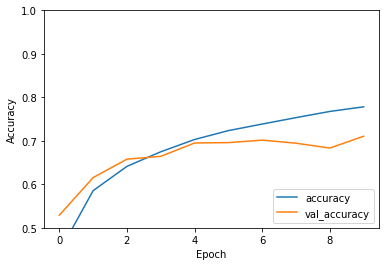


Fig 12 – Accuracy of the training and of the testing, after each epoch

The number of epochs was set to 10 and it seems to be a reasonable number, as the testing accuracy value starts to diverge from the training accuracy, indicating a good stopping point. The final test accuracy was measured as 71.05%

More analysis could be done on this model, but it was time to move on to the next architecture, which is of bigger interest.

Lastly, the CNN as proposed in the **“Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction”** paper.

The data used to train and test the network is the fashion\_mnist from the tensorflow dataset for simplification and testing purposes. The main goal now is to get a working CNN architecture that is as close as the one described on the paper as possible. This dataset is composed of 70,000 images, each having 28x28 grayscale pixels. From those images, 60,000 are used as training and 10,000 are used as testing. The dataset is equally distributed among 10 non-overlapping and distinct classes; each image has its label and will be normalized before going into the model training and testing.

As for now, the multi-scale feature extractor and the high-level representation are implemented, but due to some errors in the code, it is not working yet. The architecture is as follows:

1. The input image goes through a pooling layer with stride (1, 2) to reduce the dimension while keeping the significant features.
2. Three identical convolutional layers with filter size (3, 3) and filter number 5 – each followed by a dropout layer with rate of 0.5 – are adopted for feature extraction.
3. The results from each dropout layer are concatenated together and then fed to a convolution layer and a dropout layer that are the same as the ones before for the high-level representation.
4. The results of the final layer are flattened and presented as the output in order to do the training of the model, which was done with these parameters:

* Adam optimizer algorithm
* Mean absolute error as the loss function
* Each epoch is trained using mini-batches of 32 samples
* 200 epochs
* Two metrics are used to evaluate the model performance: mean absolute percentage error and root mean square error

The training of the flawed architecture was almost 3.5 hours long, with about one minute/epoch. The training time of the proposed model, as is, after the due corrections, are not expected to change too much.

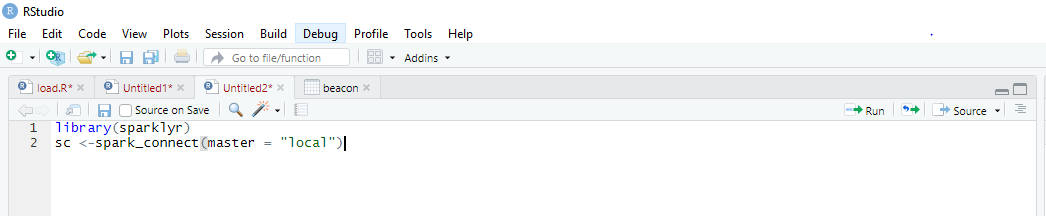
It was observed that the model might not need too much epochs to reach an optimal state. Further investigation will be carried.

**Appendix B**

**sparklyr: R interface for Apache Spark**

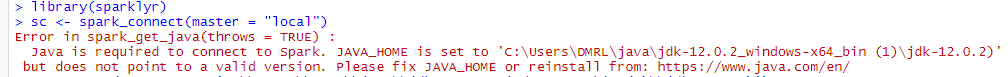
**Prerequisites:**

1. Install the sparkly package from CRAN in Rstudio.
2. Install a local version of spark for development purposes.



1. Need to connect to a local instance of Spark via the **spark\_connect** function.

**Issues Raised:**



**Issue Solved:**

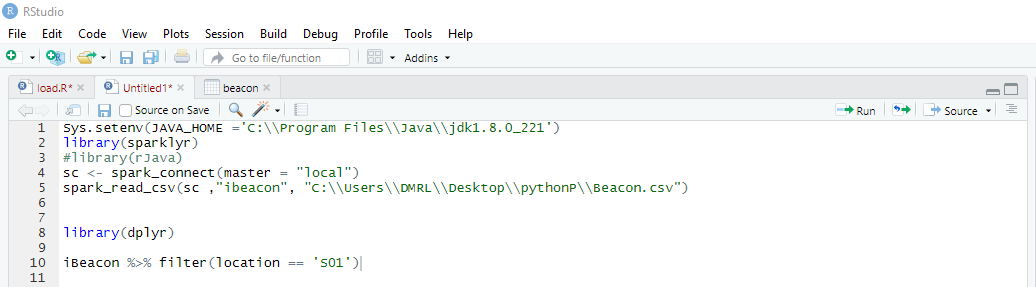
By providing the java directory in Rstudio by using Sys.setenv(JAVA\_HOME =’ java path’) and make sure Java(JDK 1.8 ) version is installed as it works for only particular version.

**Using dplyr – A grammar for Data Manipulation:**

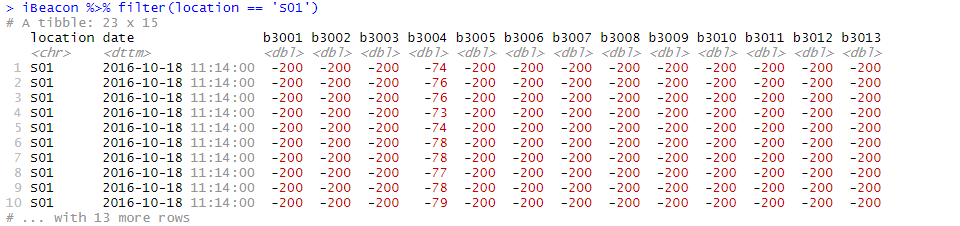
A fast and consistent toll for working with data frame like objects both in memory and out of memory.

**List of functions dplyr offers:**

* 1. **Select():Select columns from the dataset.**
  2. **Filter():filter out certain rows and columns.**

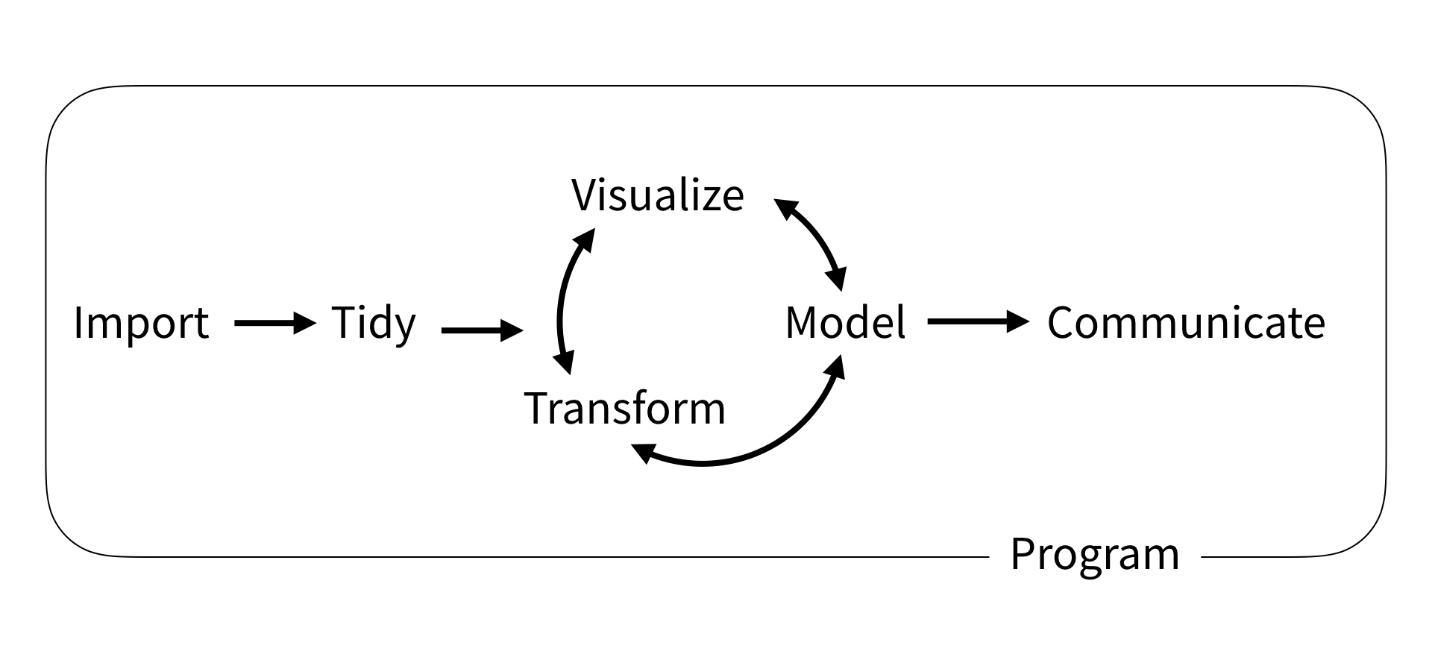


**Output:**



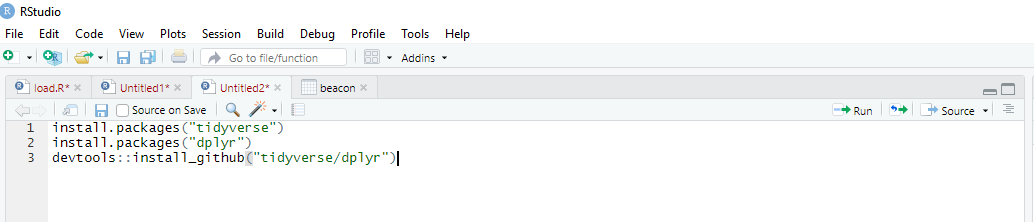
**R packages for Data Science**

**Tidyverse** is an opinionated collection of R packages designed for data science. All packages share an underlying design philosophy, grammar and data structures.

****

**Fig 1,** Structure of Tidyverse packages

**Required packages to be installed:**



**Appendix C**

**Key Accomplishments:**

1. Implemented basic, building-block code that can be scaled to multi-modal data sets.
2. Found an R-package, *densratio( )*, for density-ratio estimation using uLSIF, RuLSIF, and KLEIP methods.
3. Was able to adapt functions from *densration( )* to the multi-modal data set that Havisha provided.

**R Package for Density Ratio Estimation**

* We take two data samples *x1* and *x2* from unknown distributions *p*(*x*) and *q*(*x*) respectively, where *x1* and *x2* are *d*-dimensional real numbers.
* We then calculate the estimated density ratio, *w*(*x*):

* The package ***densratio*** provides a function *densratio()* that returns an object with a method to estimate density ratio as *compute\_density\_ratio()*.

**Implementation of *densratio()***

**Code for Plot 1:**

x1 <- rnorm(data, mean = 1, sd = 1/8)

x2 <- rnorm(data, mean = 1, sd = 1/2)

library(densratio)

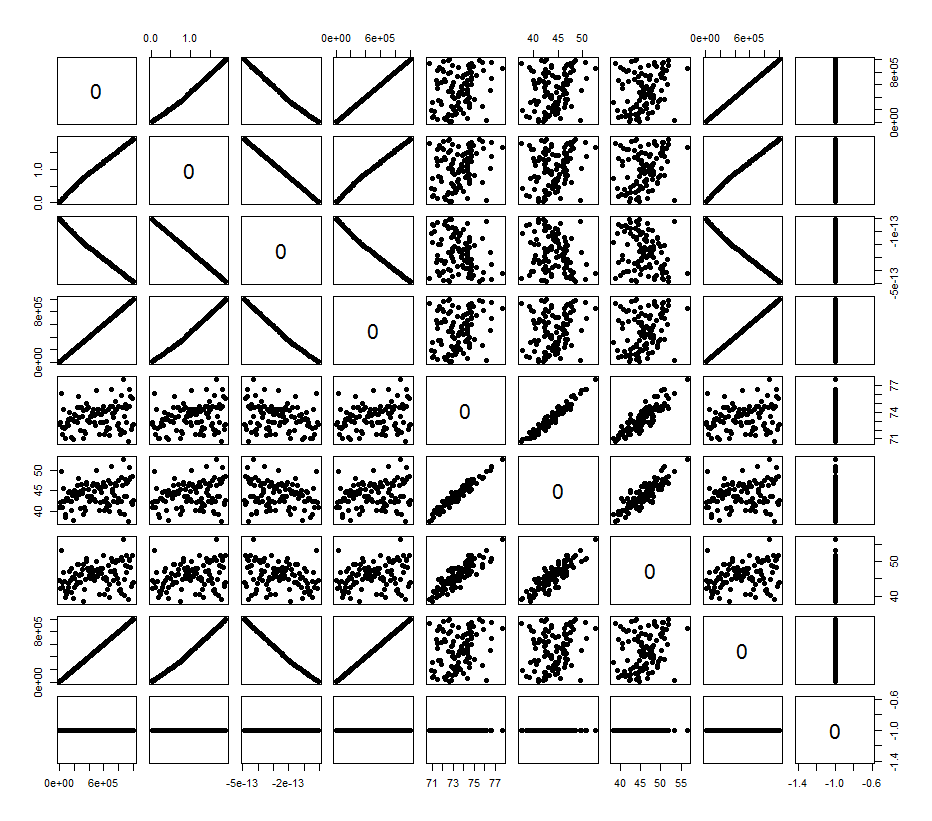
densratio\_obj <- densratio(x1, x2)

densratio\_obj

w\_hat <- densratio\_obj$compute\_density\_ratio(data)

#plot(data, w\_hat, pch=19)

**Plot 1**

Plot 1: In this case, the true density ratio is known. So we can compare with the estimated density ratio *.*

**Code for Plot 2:**

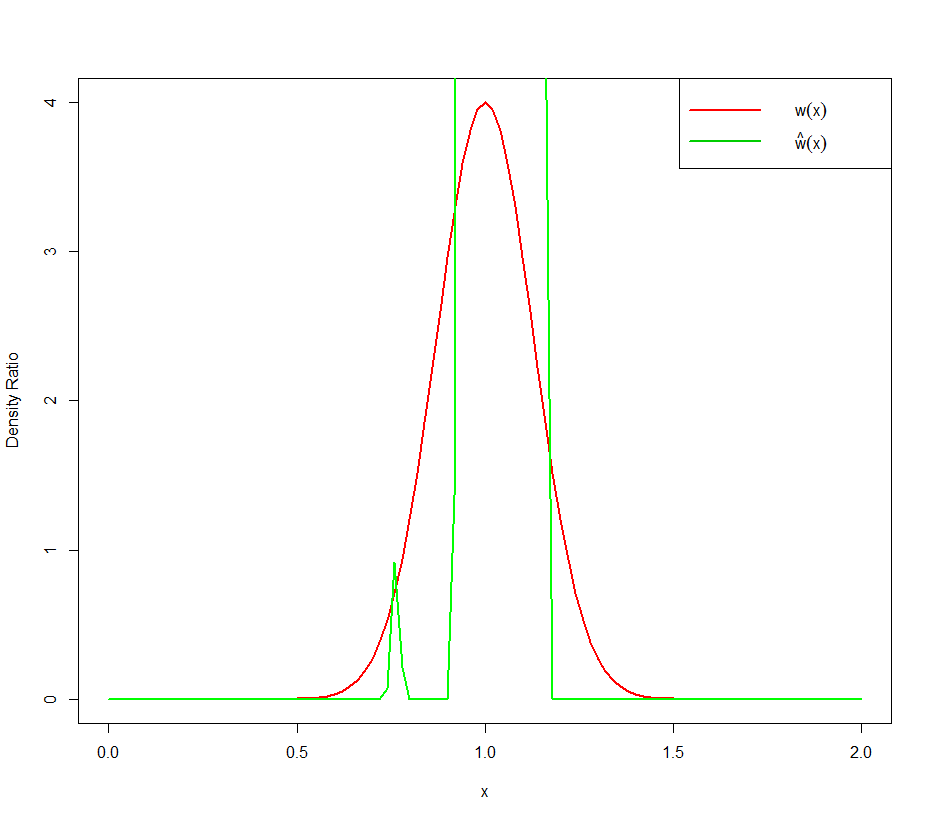
true\_density\_ratio <- function(data) dnorm(data, 1, 1/8) / dnorm(data, 1, 1/2)

plot(true\_density\_ratio, xlim=c(0, 2), lwd=2, col="red", xlab = "x", ylab = "Density Ratio")

plot(densratio\_obj$compute\_density\_ratio, xlim=c(0, 2), lwd=2, col="green", add=TRUE)

legend("topright", legend=c(expression(w(x)), expression(hat(w)(x))), col=2:3, lty=1, lwd=2, pch=NA)

**Plot 2**

****

Plot 2: The distribution of compared to the distribution of

**Appendix D**

Comparing the evaluation of NaiveBayes, HoeffdingTree and Concept Drift methods in Massive Online Analysis. The following are the results obtained.

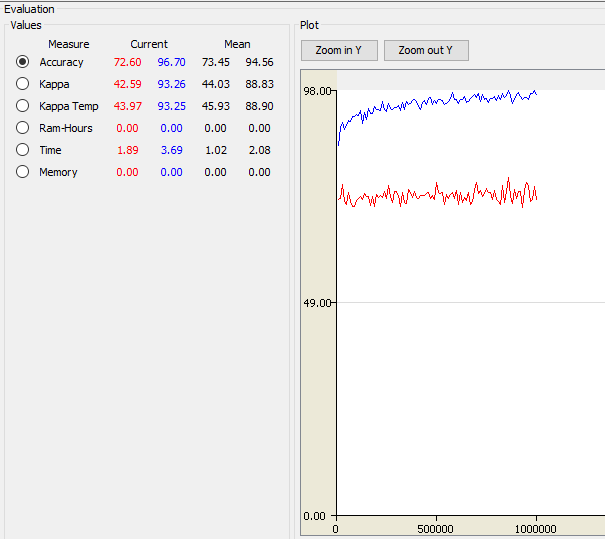


Figure 1: Comparison between the classifiers NaïveBayes and HoeffdingTree

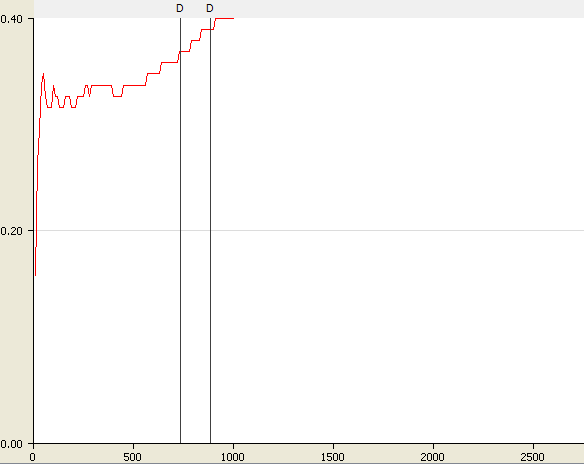


Figure 2: Evaluation of drift deduction method.

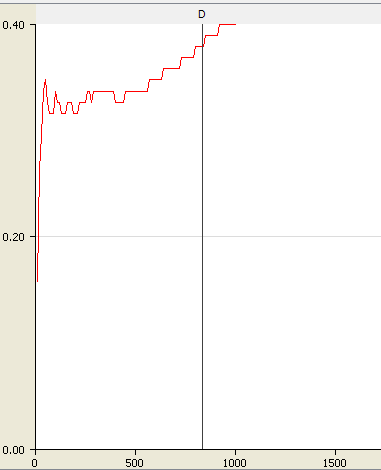


Figure 3: Evaluation of CUSUM deduction method.

1. Keras Documentation. https://keras.io/ [↑](#footnote-ref-1)
2. Xiang Li, Wei Zhang, Qian Ding (2018). Deep learning-based remaining useful life estimation of bearings using multi-scale feature extraction [↑](#footnote-ref-2)
3. <https://spark.rstudio.com/> [↑](#footnote-ref-3)
4. <https://www.tidyverse.org/> [↑](#footnote-ref-4)
5. Liu, Song & Yamada, Makoto & Collier, Nigel & Sugiyama, Masashi. (2012). Change-Point Detection in Time-Series Data by Relative Density-Ratio Estimation. Neural Networks. 10.1016/j.neunet.2013.01.012. [↑](#footnote-ref-5)
6. <https://cran.r-project.org/web/packages/densratio/vignettes/densratio.html>  [↑](#footnote-ref-6)
7. Sugiyama, M. & Suzuki, T. & Kanamori, Takafumi. (2010). Density Ratio Estimation: A Comprehensive Review. RIMS Kokyuroku. 10-31. [↑](#footnote-ref-7)
8. <https://cran.r-project.org/web/packages/RMOA/index.html> [↑](#footnote-ref-8)