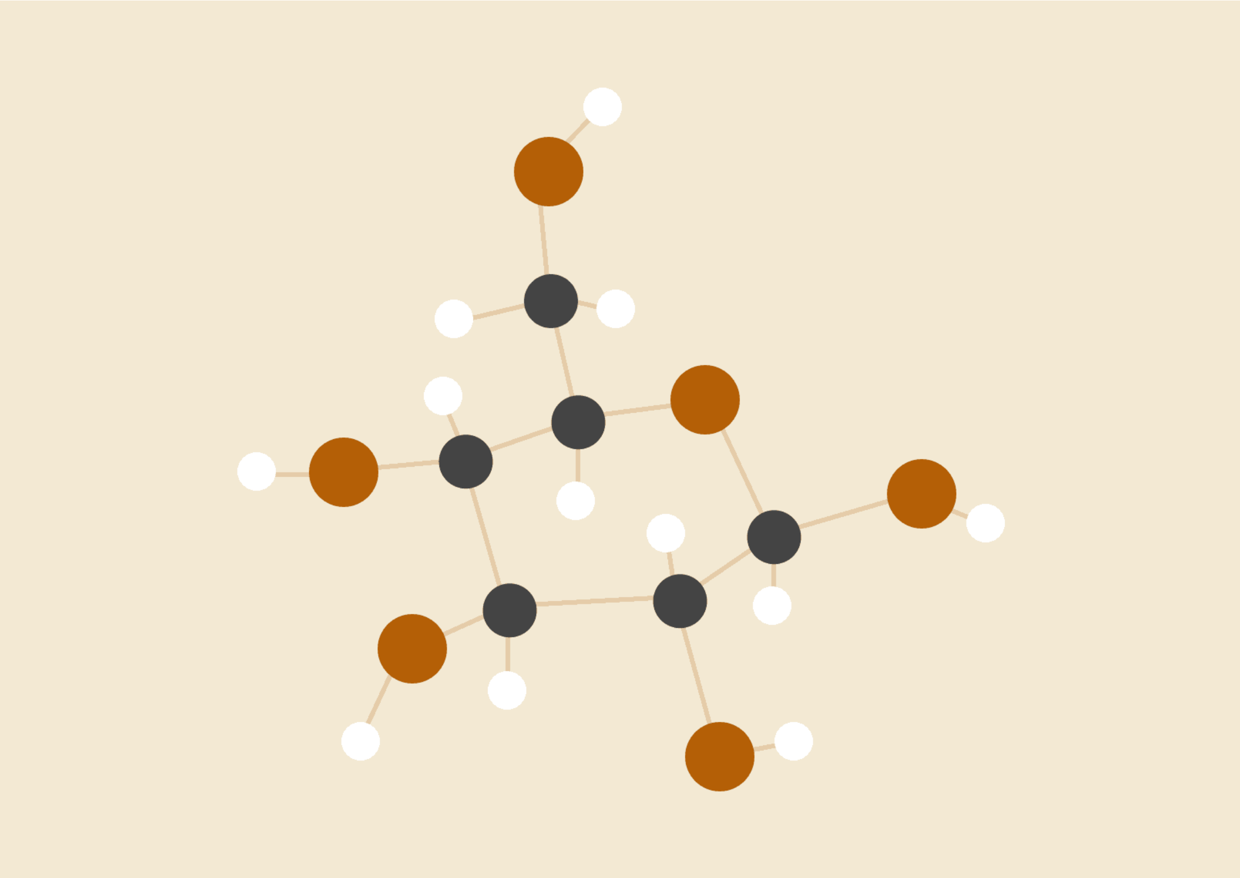
RAPPORT TP3

Cours d'apprentissage statistique de M. Dalalyan



Cordoba Muñoz Santiago
Ureña Stéphane
Watteau Rodolphe

04/12/2018
MS–DS

INTRODUCTION :

This document is intended to answer various questions related to TP3.

The document focuses on neural network techniques, more precisely it uses the Boston dataset from the library MASS, which contains a collection of data about the housing values in the suburbs, with the idea of predicting the "median value owner-occupied homes" (medv).

The Available Data :

The dataset has 506 observations for each one of the 14 variables, at the same time, there is no missing observations for any of the variables. We randomly split the data into a train set that has 75% of the observations and test set with 25%.

Since we are using some regression, we will use the mean squared error (MSE) as the measure of error (how much the predictions are far away from the real data).

Name	Description	Min	Mean	Max
Crim	Per capita crime rate by town	0.00632	3.6135	88.9762
Zn	Proportion of residential land zoned for lots over 25000 sq. ft.	0	11.36	100
Indus	Proportion of non-retail business acres per town	0.46	11.14	27.74
Chas	Charles River dummy variable (1 if tract bounds rive, 0 otherwise)	0	0.06917	1
Nox	Nitrogen oxides concentration (parts per 10 million)	0.3850	0.5547	0.8710
Rm	Average number of rooms per dwelling	3.561	6.285	8.780
Age	Proportion of owner-occupied units build prior to 1940	2.90	68.57	100
Dis	Weighted mean of distance to five Boston employment centers	1.130	3.795	12.127
Rad	Index of accessibility to radial highways	1	9.549	24
Tax	Full-value property-tax rate per $10,000	187	408.2	711
Ptratio	Pupil-teacher ratio by town	12.60	18.46	22
Black	1000/(Bn-0.63) ^2 where Bk is proportion	0.32	356.67	396.90

	of blacks by town			
Lstat	Lower status of the population (%)	1.73	12.65	37.97
Medv	Median value owner-occupied home	5	22.53	50

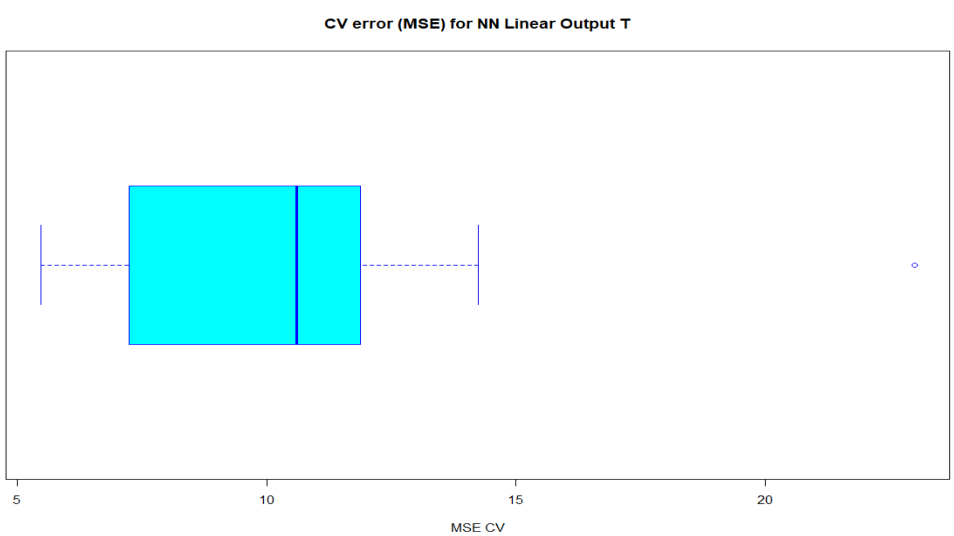
Source :

- Harrison, D. and Rubinfeld, D.L. (1978) Hedonic prices and the demand for clean air. J. Environ. Economics and Management 5, 81–102.
- Belsley D.A., Kuh, E. and Welsch, R.E. (1980) Regression Diagnostics. Identifying Influential Data and Sources of Collinearity. New York: Wiley.

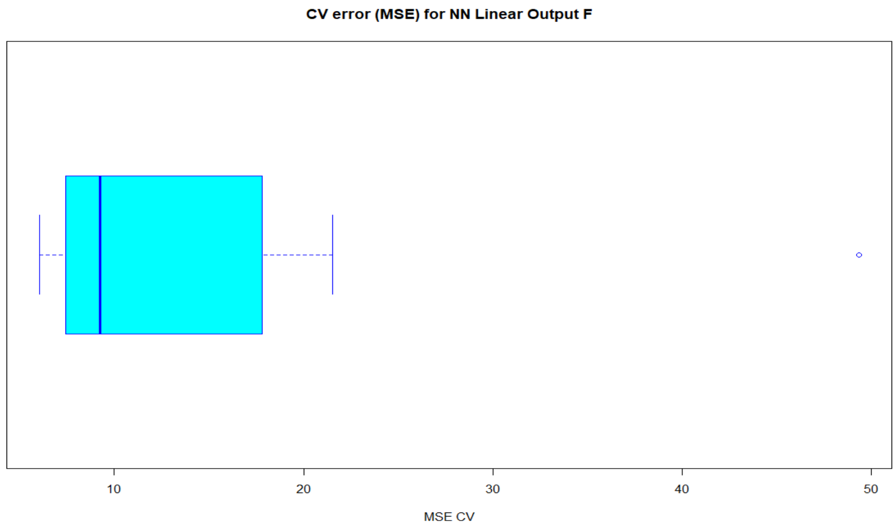
Question 1

Yes, we can. The linear output T is used so the act.fct (Activation function) is not applied to the output neurons. Act.fct is a function that is used to smooth the results of the cross product of the covariate or neurons and the weights.

In other words, linear output is used to specify whether we want to do a regression (linear output T) or a classification (linear output F).



Mean error 10.86



Mean error 15.00

As we can see, the median of the boxplot in the case of linear output =F is higher to the one in the linear output =T, but the dispersion of T is at the same time bigger than in the case of F. However, when repeated the process several times, both the median and the dispersion fluctuate for each execution, therefore, we can't say that there is a significant difference between modelling with regression or classification in terms of the dispersion and the median of the MSE.

Source : https://datascienceplus.com/fitting-neural-network-in-r/

Question 2:

Question 2.1:

One of the problems of neural-network training algorithms is that sometimes they can get stuck in a local minimum instead of finding the global optimum. Therefore, it is necessary to repeat the training a few times, with random starting values for the weights in order to find the global minimum.

The argument *rep* is the function that allows us to do so, as it is the number of repetitions for the neural networks training, in other words it allows us to repeat the training with different starting weights, so in the case of rep=10 we will get the result of 10 different neural networks

Question 2.2:

The mean square errors found using the command with rep is a bit smaller than the one without repetition. The reason is that command with repetition allows you to use the repetition that minimize the MSE (when we use the function "compute" it takes the repetition 1 by default unless we specify the contrary, so for this exercise we need to specify which repetition we want to choose).

However, both commands are applying the same procedure with random initial weights with the difference that the command using rep=10 is repeating the same process 10 times, in other words, it is applying the command without rep, 10 times.

Case	Error 1	Error 2	Error 3	Error 4	Error 5
NN without rep	9.77	7.42	9.84	8.56	13.52
NN with rep =10	6.61	7.31	7.93	6.82	7.02

Source : http://www.learnbymarketing.com/tutorials/neural-networks-in-r-tutorial/

Question 3 :

Question 3.1

The line code that allows to specify that the activation function is the hyperbolic tangent is the following.

```
nn <- neuralnet(f,data=train_,hidden=c(5,3),linear.output=T, act.fct = "tanh")
```

It is basically the same line of code that we use for the neural net but adding the "act.fct" argument.

Question 3.2

So, the tanh non-linearity squashes real numbers to range between [-1,1] so in order to apply the neural net we need to normalize and denormalize the range from the original [0,1] from the sigmoid function to the new range. The advantage of the tanh over the sigmoid it is that unlike the sigmoid neuron its output is zero-centered.

In order to scale the date to the new range we can use the min max method (the same method that is used during the TP).

$$X_{normalized} = \frac{(b-a) * x - min(x)}{max(x) - min(x)} + a$$

$$X_{normalized} = \frac{(2) * x - min(x)}{max(x) - min(x)} + 1$$

Normalization:

train_2 = as.data.frame((2*((train - min(train))/(max(train) - min(train))))-1)

test_2 = as.data.frame((2*((test - min(test))/(max(test) - min(test))))-1)

In the case of the normalization we apply the same method but on the contrary case, so we try to find the value of x.

$$x=\frac{X_{normalized}+1}{2}*(max(x)-min(x))+min\ (x)$$

Denormalization:

pr.nn_ <- (((pr.nn$net.result+1)/2)*(max(data$medv)-min(data$medv)))+min(data$medv)

Source : https://stats.stackexchange.com/questions/281162/scale-a-number-between-a-range

RÉFÉRENCES :

Question 1 :

https://datascienceplus.com/fitting-neural-network-in-r/

Question 2 :

http://www.learnbymarketing.com/tutorials/neural-networks-in-r-tutorial/

Question 3 :

https://stats.stackexchange.com/questions/281162/scale-a-number-between-a-range