Caso Práctico I

Minería de datos

Estimación y predicción con redes neuronales

José Ignacio Escribano

Móstoles, 2 de diciembre de 2015

Índice de figuras

1. Función $y = x \sin x$ con 200 puntos 1
2. Función $y = x \sin x$ con 200 puntos sin el parámetro type='l' 2
3. Recta de regresión de la función $y = x \sin x$ 3
4. Q-Q plot y gráfico de residuos del modelo lineal 3

Índice de tablas

Índice

Índice de figuras b

Índice de tablas c

1. Introducción 1

2. Resolución de las cuestiones de evaluación 1

2.1. Cuestión 1 1

2.2. Cuestión 2 1

3. Conclusiones 2

1. Introducción

En este caso práctico, pondremos en práctica lo aprendido en teoría sobre redes neuronales, a través de dos librerías distintas de R que las implementan. Éstas son `nnet` y `neuralnet`.

2. Resolución de las cuestiones de evaluación

A continuación resolveremos las cuestiones planteadas en el caso práctico. En todas las cuestiones de evaluación consideraremos la función no lineal dada por $y = x \sin x$.

2.1. Cuestión 1

Representamos la función $y = x \sin x$ tomando 200 puntos en el intervalo $[-2\pi, 2\pi]$. La función queda como se muestra en la Figura 1.

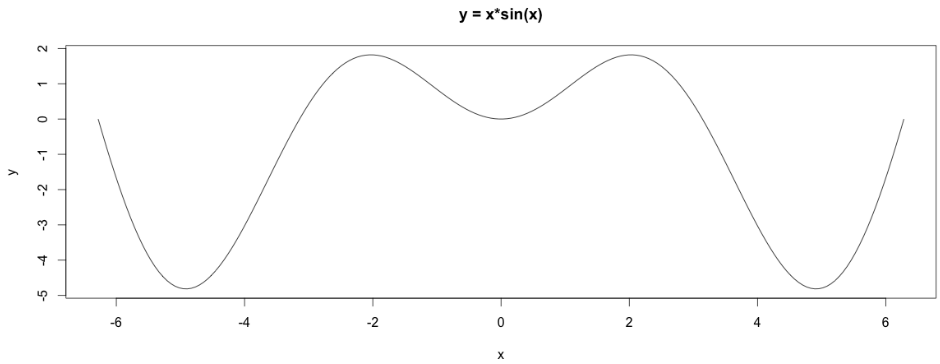


Figura 1: Función $y = x \sin x$ con 200 puntos

Si dibujamos la función sin el parámetro `type='l'`, tenemos que sólo se muestran los puntos $\{(x_i, y_i)\}$ en el plano. El parámetro `type='l'` lo que hace es interpolar los dintintos puntos para obtener una curva. La Figura 2 muestra este hecho.

Esta función sí podrá ser aproximada por una red neuronal, ya que aunque se trate de una función no lineal, el Teorema de Cybenko nos asegura que una red neuronal multicapa puede aproximar una función continua en subconjuntos compactos de $\mathbb{R}^n$, que es lo que intentamos aproximar.

2.2. Cuestión 2

Ajustamos un modelo lineal con los 200 puntos calculados anteriormente. El resumen que muestra R es el siguiente:

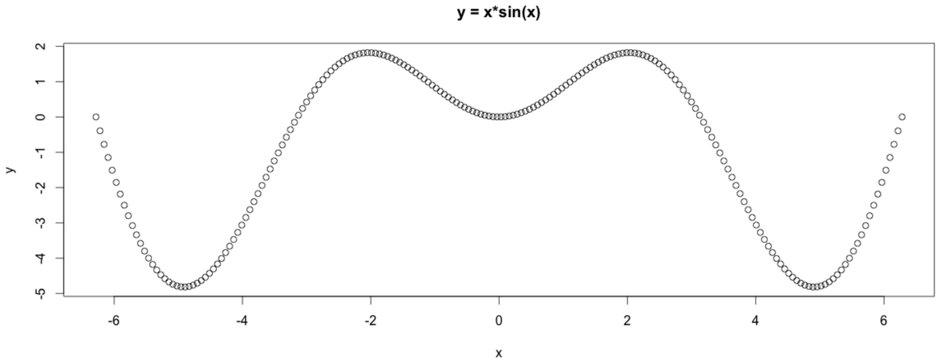


Figura 2: Función $y = x \sin x$ con 200 puntos sin el parámetro type='l'

```
Call:
lm(formula = y ~ x)

Residuals:
       Min         1Q     Median         3Q        Max
-3.8188378 -2.2907973  0.9951677  2.0555346  2.8136221

Coefficients:
              Estimate     Std. Error  t value        Pr(>|t|)
(Intercept) -0.9946693  0.16374284208 -6.07458 0.0000000062568 ***
x           -0.0000000  0.04491295114  0.00000               1
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.315673 on 198 degrees of freedom
Multiple R-squared:  6.660172e-32,Adjusted R-squared:  -0.005050505
F-statistic: 1.318714e-29 on 1 and 198 DF,  p-value: 1
```

La recta de regresión es $y = -1$. La Figura 3 muestra la recta de regresión sobre la función. La Figura indica un ajuste bastante malo. Esto lo corroboramos con el coeficiente de determinación, que es prácticamente 0, por lo que no hay relación lineal entre los valores de x e y.

Comprobemos que se incumplen las hipótesis del modelo: normalidad en los residuos y/o linealidad. Para ello, representamos en R, un Q-Q plot y un gráfico de residuos (Figura 4).

3. Conclusiones

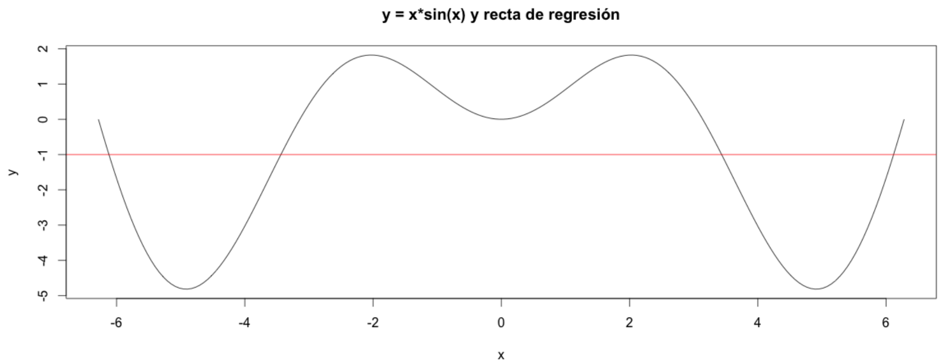
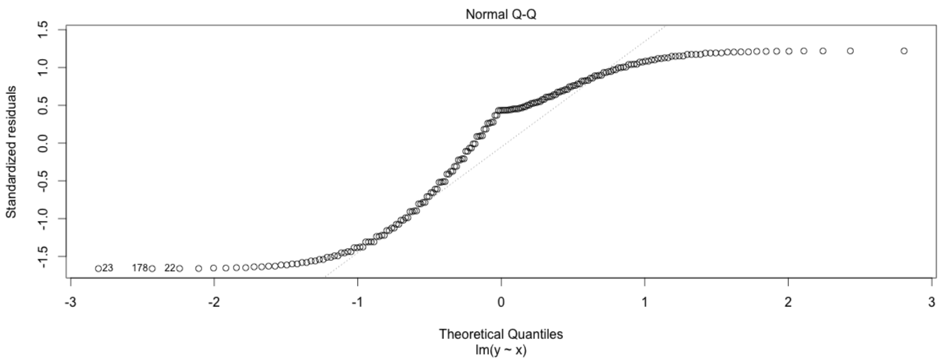
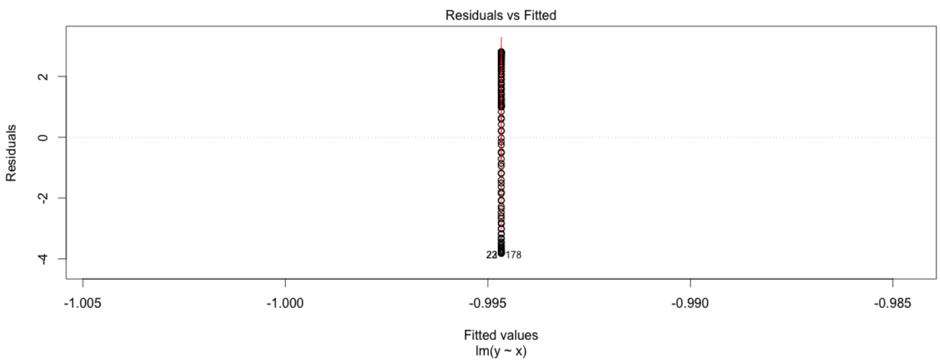


Figura 3: Recta de regresión de la función $y = x \sin x$



(a) Q-Q plot



(b) Residuos

Figura 4: Q-Q plot y gráfico de residuos del modelo lineal