

Práctica 5

L1-Predicción

Luis Eduardo Galindo Amaya (1274895)



Asignatura	Estadística Avanzada
Docente	Olivia Mendoza Duarte
Fecha	22-09-2022

L1-Predicción

Luis Eduardo Galindo Amaya (1274895)

22-09-2022

Instrucciones

En ésta práctica usarás como base el ejemplo de regresión lineal simple en R visto en clase, para resolver un problema de predicción, mismo que propondrás del sitio del repositorio de la UCI (elegir un dataset distinto de las anteriores prácticas).

Entregarás el archivo con el código en R, capturas de pantalla de los resultados y una explicación del problema y la solución.

Dataset

Breast Cancer Wisconsin (Diagnostic)

- 1. ID number
- 2. Diagnosis (M = malignant, B = benign)

Ten real-valued features are computed for each cell nucleus:

- a. radius (mean of distances from center to points on the perimeter)
- b. texture (standard deviation of gray-scale values)
- c. perimeter
- d. area
- e. smoothness (local variation in radius lengths)

- f. compactness ($\text{perimeter}^2 / \text{area} - 1.0$)
- g. concavity (severity of concave portions of the contour)
- h. concave points (number of concave portions of the contour)
- i. symmetry
- j. fractal dimension (coastline approximation 1)

Origen

Descripcion del problema

Se espera encontrar una relacion entre el tamaño del tumor y su radio, los datos del dataset son compuestas de multiples mediciones. por lo que son un aproximado del tamaño del tumor.

Código R

```
archivo = read.csv("./dataset/wdbc.data", nrow = 30)
x = archivo[3] #radio
y = archivo[5] #perimetro

datos = data.frame(x,y)
datos
regresion = lm(datos)

b= regresion$coefficients[1]
m = regresion$coefficients[2]
b
m

yr = m * x + b
yr
datos_r = data.frame(x,yr)
datos_r

jpeg(file="img/rp.jpeg")
par(mfrow = c(2,1))
plot(datos)
plot(datos_r, type="l")
dev.off()
```

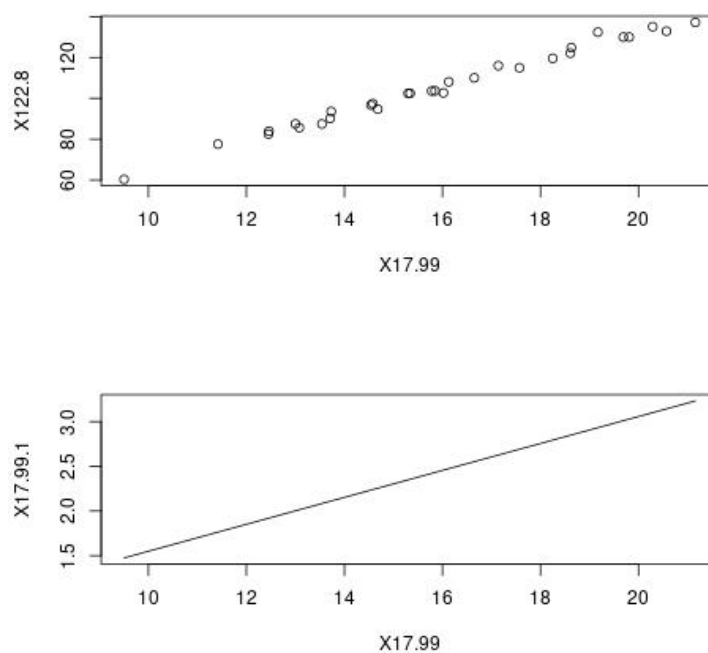


Figura 1: 'y' Radio del tumor, 'x' perimetro.

Regresion resultante

$$0,1441553x - 0,807$$

	x2	y _{pred}
1	1	0.9525883
2	2	1.0972761
3	3	1.2419639
4	4	1.3866517
5	5	1.5313394
6	6	1.6760272
23	23	4.1357197
24	24	4.2804075
25	25	4.4250953

```
archivo = read.csv("../dataset/wdbc.data")
```

```
x = archivo[1:300,3] #radio
```

```
y = archivo[1:300,5] #perimetro
```

```
datos = data.frame(x,y)
```

```
regresion = lm(datos)
```

```
b = regresion$coefficients[1]
m = regresion$coefficients[2]
yr = m * x + b

datos_r = data.frame(x,yr)
datos_r

## x2 = archivo[30:100,3] #radio
x2 = 1:300
y_pred = m * x2 + b

jpeg(file="img/rp2.jpeg")
par(mfrow = c(3,1))
plot(datos)
plot(datos_r, type="l")
plot(y_pred, type="l")
dev.off()
```

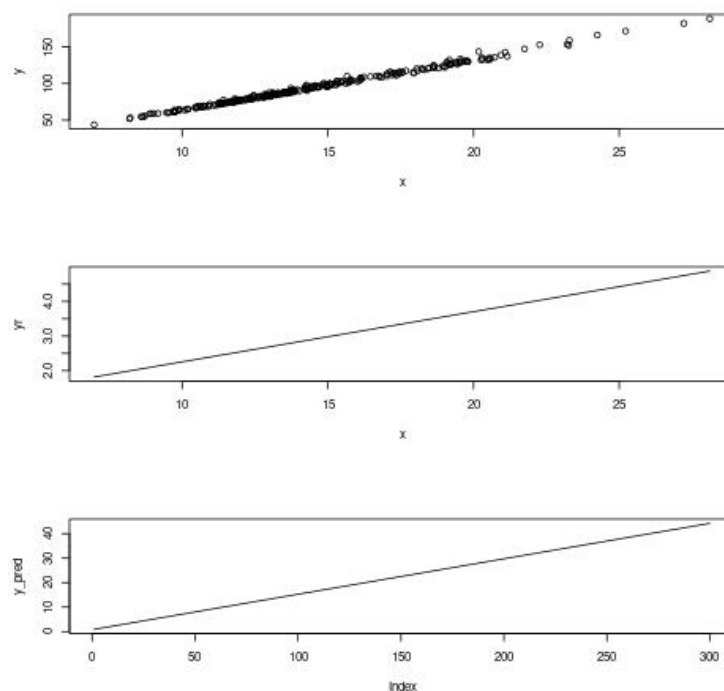


Figura 2: 'y' Radio del tumor, 'x' perimetro