

Regresión Polinómica

jerf

15/6/2021

Regresión Polinómica

```
dataset = read.csv("Position_Salaries.csv")
dataset = dataset[, 2:3]
dataset
```

```
##      Level  Salary
## 1         1   45000
## 2         2   50000
## 3         3   60000
## 4         4   80000
## 5         5  110000
## 6         6  150000
## 7         7  200000
## 8         8  300000
## 9         9  500000
## 10        10 1000000
```

Pre Procesado de datos

No es necesario dividir el conjunto de datos, pues son muy pocos

Escalado de datos

Tampoco es necesario en este caso

Ajustar Mdelo de Regresión Lineal con el dataset

```
lin_reg = lm(formula = Salary ~ Level,
              data = dataset
              )
summary(lin_reg)
```

```
##
## Call:
## lm(formula = Salary ~ Level, data = dataset)
##
## Residuals:
```

	Min	1Q	Median	3Q	Max
##	-170818	-129720	-40379	65856	386545

```
##
```

```
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) -195333    124790  -1.565  0.15615
## Level       80879      20112   4.021  0.00383 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 182700 on 8 degrees of freedom
## Multiple R-squared:  0.669, Adjusted R-squared:  0.6277
## F-statistic: 16.17 on 1 and 8 DF,  p-value: 0.003833
```

Un empleado en nivel cero debería pagar 195333 a la empresa, por cada nivel que suba el empleado ganaría 80879 más

Ajustar Modelo de Regresión Polinómica con el dataset

```
dataset$Level2 = dataset$Level ^ 2
dataset$Level3 = dataset$Level ^ 3
dataset$Level4 = dataset$Level ^ 4
poly_reg = lm(formula = Salary ~ Level + Level2 + Level3 + Level4,
              data = dataset)

summary(poly_reg)

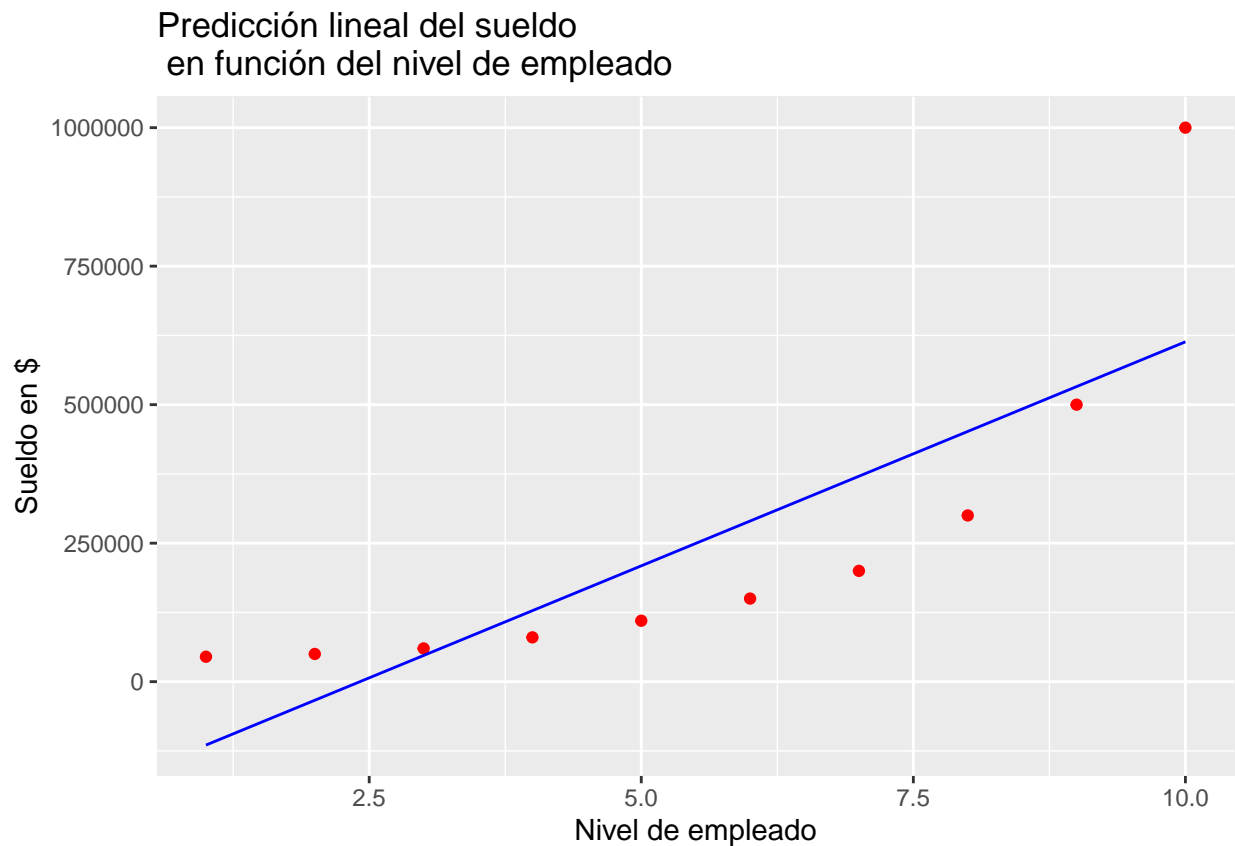
##
## Call:
## lm(formula = Salary ~ Level + Level2 + Level3 + Level4, data = dataset)
##
## Residuals:
##      1      2      3      4      5      6      7      8      9     10
## -8357 18240  1358 -14633 -11725   6725  15997  10006 -28695  11084
##
## Coefficients:
##           Estimate Std. Error t value Pr(>|t|)
## (Intercept) 184166.7    67768.0   2.718  0.04189 *
## Level      -211002.3    76382.2  -2.762  0.03972 *
## Level2       94765.4    26454.2   3.582  0.01584 *
## Level3     -15463.3     3535.0  -4.374  0.00719 **
## Level4        890.2      159.8   5.570  0.00257 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20510 on 5 degrees of freedom
## Multiple R-squared:  0.9974, Adjusted R-squared:  0.9953
## F-statistic: 478.1 on 4 and 5 DF,  p-value: 1.213e-06
```

Visualización del modelo Lineal

```
library(ggplot2)

ggplot() +
  geom_point(aes(x = dataset$Level, y = dataset$Salary),
            colour = "red") +
```

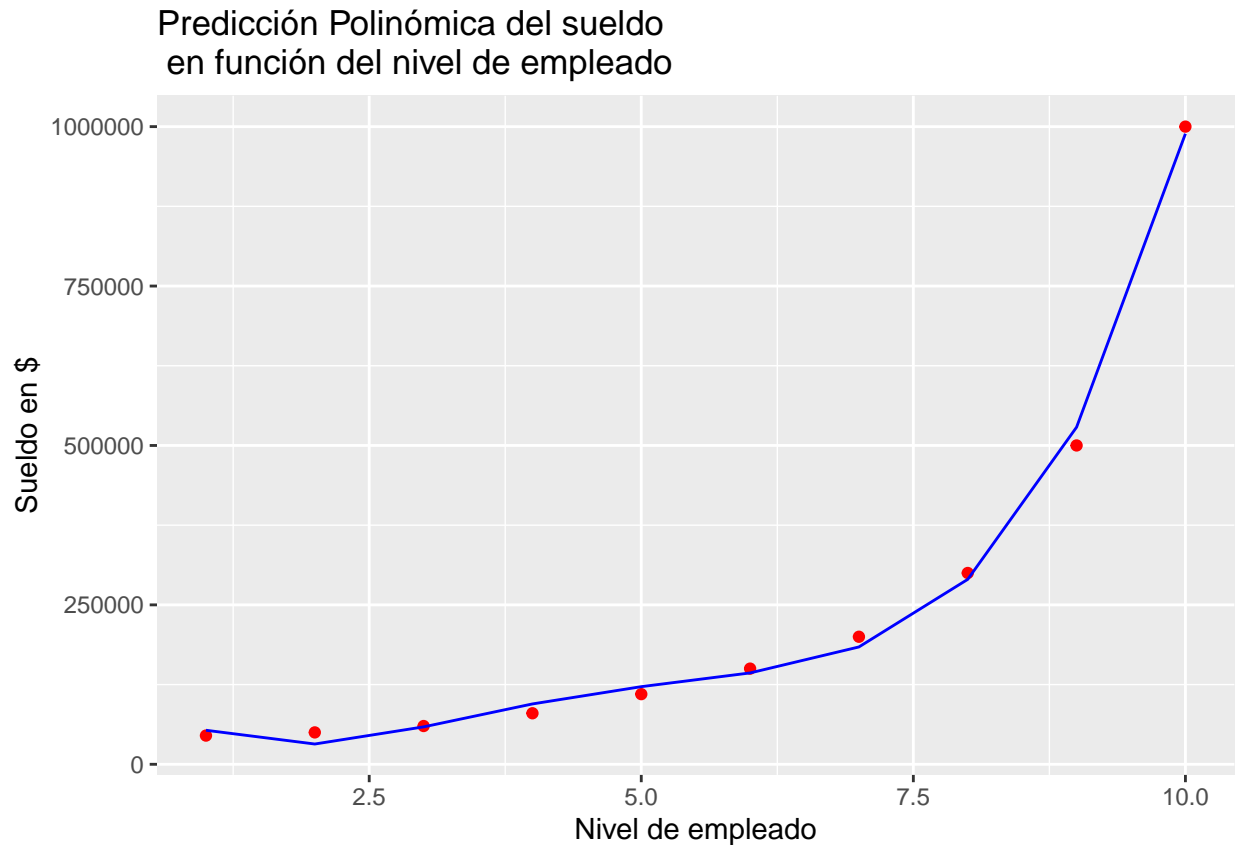
```
geom_line(aes(x = dataset$Level, y = predict(lin_reg, newdata = dataset)),
          colour = "blue")+
xlab("Nivel de empleado") +
ylab("Sueldo en $") +
ggtitle("Predicción lineal del sueldo \n en función del nivel de empleado")
```



Visualización del modelo Polinómico

```
library(ggplot2)

ggplot() +
  geom_point(aes(x = dataset$Level, y = dataset$Salary),
            colour = "red") +
  geom_line(aes(x = dataset$Level, y = predict(poly_reg, newdata = dataset)),
            colour = "blue")+
  xlab("Nivel de empleado") +
  ylab("Sueldo en $") +
  ggtitle("Predicción Polinómica del sueldo \n en función del nivel de empleado")
```



Predicción de resultados con Regresión Lineal

```
y_pred = predict(lin_reg, newdata = data.frame(Level = 6.5))
y_pred
```

```
##          1
## 330378.8
```

Es una cantidad mucho más elevada de la que se considera en los datos

Predicción de resultados con Regresión Polinómica

```
y_pred = predict(poly_reg, newdata = data.frame(Level = 6.5 , Level2 = 6.5 ^2, Level3 = 6.5 ^3, Level4 = 6.5 ^4))
y_pred
```

```
##          1
## 158862.5
```

La predicción es mucho más acertada, por lo cual este modelo trabaja mejor