

Regresion Lineal Simple

Diego Pavez

15/6/2020

```
# Importar el dataset
dataset = read.csv('Salary_Data.csv')
head(dataset)
```

```
##   YearsExperience Salary
## 1             1.1  39343
## 2             1.3  46205
## 3             1.5  37731
## 4             2.0  43525
## 5             2.2  39891
## 6             2.9  56642
```

```
# Dividir los datos en conjunto de entrenamiento y conjunto de test
# install.packages("caTools")
library(caTools)
set.seed(123)
split = sample.split(dataset$Salary, SplitRatio = 2/3)
training_set = subset(dataset, split == TRUE)
testing_set = subset(dataset, split == FALSE)
```

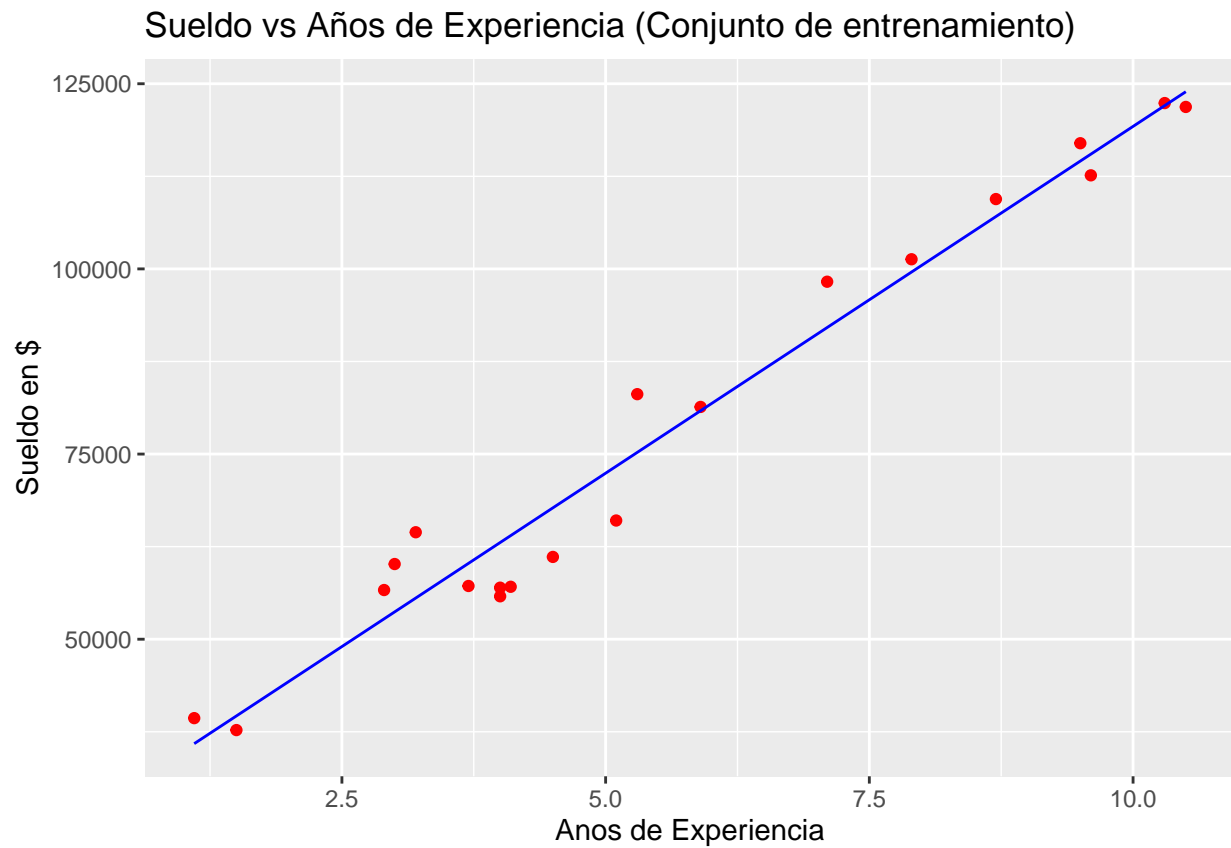
```
# Ajustar el modelo de regresión lineal simple con el conjunto de entrenamiento
regressor = lm(formula = Salary ~ YearsExperience,
               data = training_set)
```

```
# Predecir resultados con el conjunto de test
y_pred = predict(regressor, newdata = testing_set)
head(y_pred, 10)
```

```
##           2           4           5           8          11          16          20          21
##  37766.77  44322.33  46195.35  55560.43  62115.99  71481.07  81782.66  89274.72
##           24          26
## 102385.84 109877.90
```

```
library(ggplot2)
ggplot() +
  geom_point(aes(x = training_set$YearsExperience, y = training_set$Salary),
            colour = "red") +
  geom_line(aes(x = training_set$YearsExperience,
               y = predict(regressor, newdata = training_set)),
            colour = "blue") +
```

```
ggtitle("Sueldo vs Años de Experiencia (Conjunto de entrenamiento)") +
xlab("Años de Experiencia") +
ylab("Sueldo en $")
```



```
library(ggplot2)
ggplot() +
  geom_point(aes(x = testing_set$YearsExperience, y = testing_set$Salary),
    colour = "red") +
  geom_line(aes(x = training_set$YearsExperience,
    y = predict(regressor, newdata = training_set)),
    colour = "blue") +
  ggtitle("Sueldo vs Años de Experiencia (Conjunto de testing)") +
  xlab("Años de Experiencia") +
  ylab("Sueldo en $")
```

