Untitled

Sergio Arnold

23/3/2021

```r
## REGRESION LINEAL SIMPLE
# Importar el dataset
dataset = read.csv('/home/sergio/Documentos/GitHub/machinelearning-az/datasets/Part 2 - Regression/Secti
#dataset = dataset[, 2:3]

# 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)

# Escalado de valores
# training_set[,2:3] = scale(training_set[,2:3])
# testing_set[,2:3] = scale(testing_set[,2:3])

regressor = lm(formula = Salary ~ YearsExperience,
               data = training_set)
summary(regressor)
```

```
##
## Call:
## lm(formula = Salary ~ YearsExperience, data = training_set)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -7325.1 -3814.4   427.7  3559.7  8884.6
##
## Coefficients:
##                 Estimate Std. Error t value Pr(>|t|)
## (Intercept)        25592       2646   9.672 1.49e-08 ***
## YearsExperience     9365        421  22.245 1.52e-14 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5391 on 18 degrees of freedom
## Multiple R-squared:  0.9649, Adjusted R-squared:  0.963
## F-statistic: 494.8 on 1 and 18 DF,  p-value: 1.524e-14
```

```r
## PREDECIR RESULTADOS CON EL CONJUNTO DE TEST

y_pred = predict(regressor, newdata = testing_set)
```

```
## VICUALIZACION DE LOS RESULTADOS EN EL CONJUNTO DE ENTRENAMIENTO

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 Año de Experiencja (Conjunto de Entrenamiento)") +
  xlab("Años de Experiencia") +
  ylab("Sueldo (en $)")
```

Sueldo Vs Año de Experiencja (Conjunto de Entrenamiento)
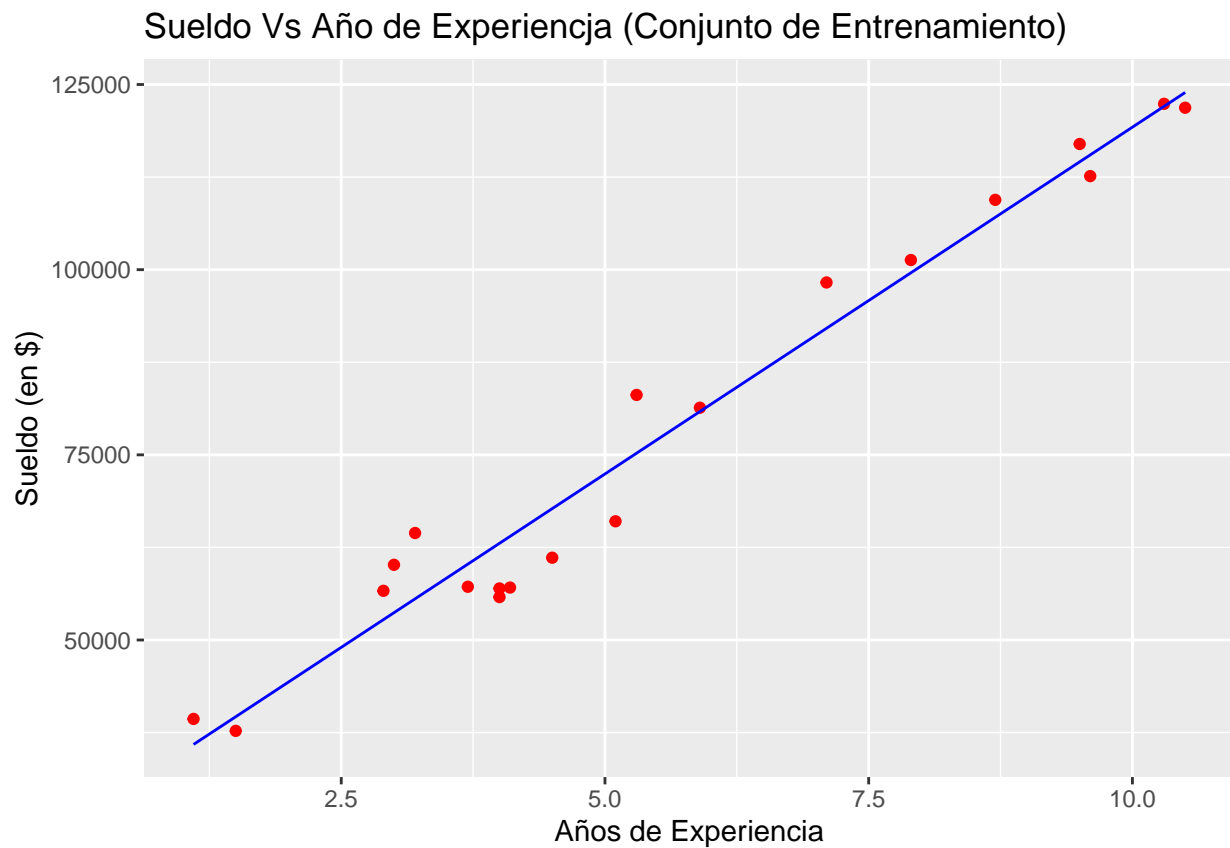

```
## VICUALIZACION DE LOS RESULTADOS EN EL CONJUNTO DE testing

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 Año de Experiencja (Conjunto de testing)") +
  xlab("Años de Experiencia") +
  ylab("Sueldo (en $)")
```

Sueldo Vs Año de Experiencja (Conjunto de testing)