

1 MACHINE LEARNING – REGRESIÓN LOGÍSTICA

Los ejemplos vistos hasta ahora incluyen un tipo de predicción de valores continuos, valor de casas, predicción de temperatura, etc.

Pero cuando queremos realizar predicciones del tipo:

- EMAIL es SPAM o no
- Comprará un cliente un determinado seguro?
- A qué partido político votará una persona?

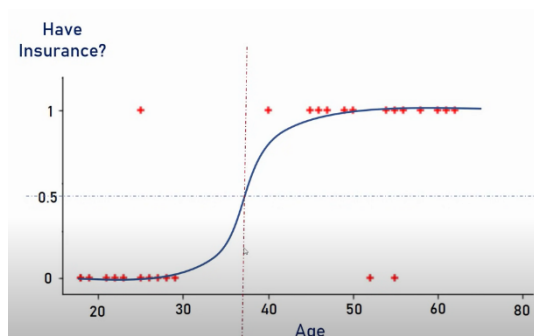
En estos problemas, la variable a predecir es CATEGÓRICA ya que lo que se intenta saber es un valor concreto de unas categorías determinadas (SI, NO , ... etc).

Este tipo de problemas se denominan de clasificación y uno de los métodos para solucionarlos es el denominado regresión logística.

Cuando la solución es entre dos valores (SI,NO) por ejemplo, se denomina CLASIFICACIÓN BINARIA y cuando hay más valores posibles se denomina clasificación MULTICLASE.

Imaginemos un dataset con edades de personas y si tienen seguro de vida o no.

age	have_insurance
22	0
25	0
47	1
52	0
46	1
56	1
55	0
60	1
62	1
61	1
18	0
28	0
27	0
29	0
49	1



Como vemos en el gráfico, la línea resultante no es una línea recta como una regresión lineal, sino un SIGMOIDE o función lógica



GENERALITAT
VALENCIANA



UNIÓN EUROPEA
Fondo Social Europeo
El FSE invierte en tu futuro

03008915 C/ Ferrocarril, 22, 03570 La Vila Joiosa Tel 966870140 Fax 966870141 <http://portal.edu.gva.es/iesmarcoszaragoza>

Primero se divide el conjunto de datos en entrenamiento y test.

```
X_train, x_test, y_train, y_test =  
train_test_split(df['age'], df.bought_insurance, test_size=0.1)
```

Y se entrena el modelo de regression logística

```
from sklearn.linear_model import LogisticRegression  
X_train, X_test, y_train, y_test = train_test_split(X,y,test_size=0.3)  
Model = LogisticRegression()  
  
Model.fit(X_train, y_train)
```

Finalmente se predicen los datos y se verifica el resultado.

```
Model.predict(X_test)  
Model.score(X_test,y_test)
```