

Módulo 5  
Clase 4

# Fundamentos de Machine Learning



Variables Categóricas

# Variables Categóricas en Regresiones

A veces necesitamos incorporar variables categóricas a nuestro modelo regresivo.

	Gender	Height	Weight
0	Male	73.847017	241.893563
1	Male	68.781904	162.310473
2	Male	74.110105	212.740856
3	Male	71.730978	220.042470
4	Male	69.881796	206.349801



# Variables Categóricas en Regresiones

Para esto, debemos pivotar los valores categóricos en las columnas y utilizar los valores 0 y 1 para indicar su valor en cada registro. Para esto, la función `get_dummies()` nos ayuda a realizar esta transformación.

```
pd.get_dummies(ds)
```

	Height	Weight	Gender_Female	Gender_Male
0	73.847017	241.893563	0	1
1	68.781904	162.310473	0	1
2	74.110105	212.740856	0	1
3	71.730978	220.042470	0	1
4	69.881796	206.349801	0	1
5	67.253016	152.212156	0	1
6	68.785081	183.927889	0	1
7	68.348516	167.971110	0	1
8	67.018950	175.929440	0	1

# Variables Categóricas en Regresiones

No obstante, estas dos nuevas columnas agregadas son multicolineales, es decir, una depende de la otra. Por eso debemos eliminar una de las columnas. Para esto, utilizamos el parámetro

**drop\_first=True.**

```
pd.get_dummies(ds, drop_first=True)
```

	Height	Weight	Gender_Male
0	73.847017	241.893563	1
1	68.781904	162.310473	1
2	74.110105	212.740856	1
3	71.730978	220.042470	1
4	69.881796	206.349801	1
5	67.253016	152.212156	1
6	68.785081	183.927889	1
7	68.348516	167.971110	1
8	67.018950	175.929440	1
9	63.456494	156.399676	1
10	71.195382	186.604926	1
11	71.640805	213.741169	1



Gracias