

Análisis de Datos y Aprendizaje Máquina con Tensorflow 2.0: Clasificación

2019/09/30

1 Regresión Logística

Objetivo: Se pondrán en práctica las herramientas para pre-procesamiento de datos. Obtener mínimo un 82% de precisión y matriz de confusión.

- Documentación: https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.LogisticRegression.html

Tiempo máximo: 2 horas

- Nota: Pre-procesar los datos para obtener un error menor

La regresión logística predice la probabilidad de que un elemento pertenezca a la clase 0 ó 1 aplicando la función sigmoide a una función lineal

$$y(x) = g(w^T x + b)$$

donde $g(z)$ es la función sigmoide

$$\sigma(z) = \frac{1}{1 + e^{-z}}$$

1.1 Importar bibliotecas y dataset

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
%matplotlib inline

data = pd.read_csv('titanic_train.csv') # Dataset de entrenamiento
data.head()
```

```
Out[1]:
```

	PassengerId	Survived	Pclass	\
0	1	0	3	
1	2	1	1	
2	3	1	3	
3	4	1	1	
4	5	0	3	

	Name	Sex	Age	SibSp	\
0	Braund, Mr. Owen Harris	male	22.0	1	
1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	
2	Heikkinen, Miss. Laina	female	26.0	0	
3	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	
4	Allen, Mr. William Henry	male	35.0	0	

	Parch	Ticket	Fare	Cabin	Embarked
0	0	A/5 21171	7.2500	NaN	S
1	0	PC 17599	71.2833	C85	C
2	0	STON/O2. 3101282	7.9250	NaN	S
3	0	113803	53.1000	C123	S
4	0	373450	8.0500	NaN	S