

# Tarea2\_SpamFilter

December 14, 2017

## 1 Tarea 2. Spam Filter.

L.E. Rojón  
138442

```
In [2]: import numpy as np
import pandas as pd
from sklearn.model_selection import train_test_split
import math
from sklearn.metrics import confusion_matrix
from scipy.stats import norm
from sklearn import preprocessing
from random import random
```

```
In [35]: df=pd.read_table('https://archive.ics.uci.edu/ml/machine-learning-databases/spambase/spambase1.txt',
header=None)
df.head()
```

```
Out[35]:
```

	0	1	2	3	4	5	6	7	8	9	...	48	49	\
0	0.00	0.64	0.64	0.0	0.32	0.00	0.00	0.00	0.00	0.00	...	0.00	0.000	
1	0.21	0.28	0.50	0.0	0.14	0.28	0.21	0.07	0.00	0.94	...	0.00	0.132	
2	0.06	0.00	0.71	0.0	1.23	0.19	0.19	0.12	0.64	0.25	...	0.01	0.143	
3	0.00	0.00	0.00	0.0	0.63	0.00	0.31	0.63	0.31	0.63	...	0.00	0.137	
4	0.00	0.00	0.00	0.0	0.63	0.00	0.31	0.63	0.31	0.63	...	0.00	0.135	

  

	50	51	52	53	54	55	56	57
0	0.0	0.778	0.000	0.000	3.756	61	278	1
1	0.0	0.372	0.180	0.048	5.114	101	1028	1
2	0.0	0.276	0.184	0.010	9.821	485	2259	1
3	0.0	0.137	0.000	0.000	3.537	40	191	1
4	0.0	0.135	0.000	0.000	3.537	40	191	1

[5 rows x 58 columns]

Hacemos una partición de los datos para entrenar con el 70% y después evaluar el modelo con el otro 30%.