

# TP6-2

## FILTRO DE SPAM



---

Asignatura: Inteligencia Artificial  
Fecha: 15/01/2021  
Autor: Diego Marco Beisty, 755232

## A) Comparación Naive Bayes con distribución Multinomial y con Bernuilli

### Distribución multinomial

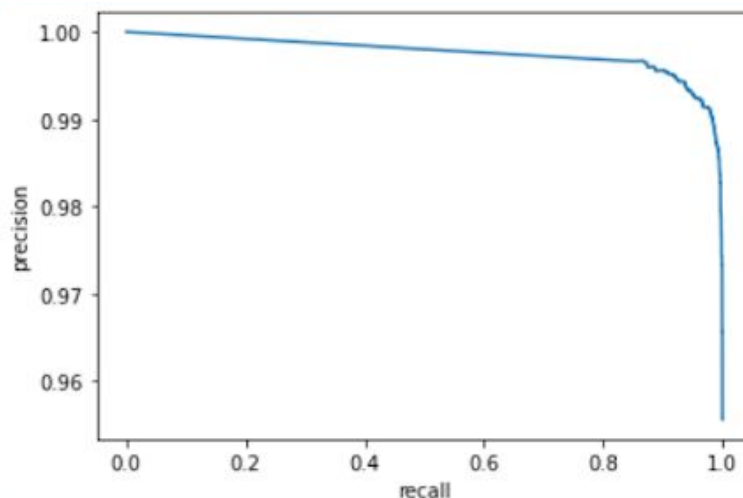
Para entrenar un clasificador con la distribución multinomial se ha hecho uso de la clase MultinomialNB.

Mediante el algoritmo K-fold cross-validation se han medido los resultados de los datos de entrenamiento para cada valor del hiper parámetro de Laplace comprendido entre 0,1 y 10,0.

Se ha decidido utilizar la métrica f1\_score para medir los resultados.

Se han obtenidos los siguientes resultados:

```
RESULTADOS
-----
Mejor hiper parámetro suavizado Laplace: 0.1
f1-score: 0.9897021370833795
matriz de confusión:
[[1438  62]
 [ 31 4469]]
curva precision-recall:
```

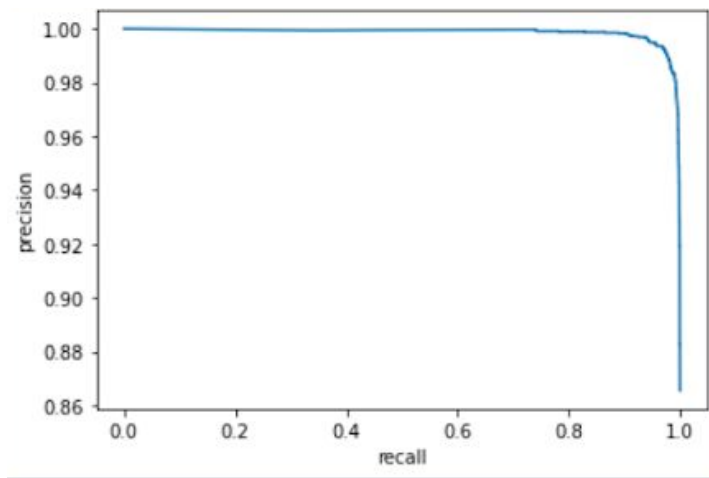


### Distribución Bernuilli

Para entrenar un clasificador con la distribución bernuilli se ha hecho uso de la clase BernuilliNB. Se han evaluado los resultados con la misma métrica (f1\_score) para el mismo rango de valores del hiper parámetro de Laplace.

Se han obtenido los siguientes resultados:

```
RESULTADOS
-----
Mejor hiper parámetro suavizado Laplace: 0.1
f1-score: 0.9153725773444084
matriz de confusión:
[[1495   5]
 [ 698 3002]]
curva precision-recall:
```



Se puede concluir que para ambas distribuciones, el valor de Laplace que mejores resultados obtiene es el 0,1.

Si nos fijamos en la métrica `f1_score`, ambas distribuciones están por encima de 0.9, sin embargo la distribución multinomial es superior (0.9897), por lo tanto se puede deducir que funciona mejor que la distribución de bernuilli.

## B) Observación de distintos valores del hiper parámetro del suavizado de Laplace

Puesto que la distribución multinomial ha generado un mejor resultado, se han obtenido el resto de valores de `f1_score` al variar el hiper parámetro de Laplace con valores inferiores y superiores a 1.

```
size=0.01, f1_score=0.9899235965009412
size=0.1, f1_score=0.9897021370833795
size=0.5, f1_score=0.9892520775623269
size=1, f1_score=0.9886792452830189
size=5, f1_score=0.9857015192135836
size=10, f1_score=0.9751187514137072
size=15, f1_score=0.9592822636300897
size=20, f1_score=0.9408457303502401
size=50, f1_score=0.8398713826366558
size=80, f1_score=0.7727272727272726
size=100, f1_score=0.7238499646142958
```

Se puede observar que con valores de “size” inferiores a la unidad se obtienen resultados muy elevados. Con 0.01, 0.1 y 0.5 el valor de `f1_score` no disminuye de 0.989.

Sin embargo, cuando el valor de “size” es igual o superior a la unidad, el valor de `f1_score` disminuye notablemente a valores entre 0.988 y 0.72.

Esto sucede porque contra mayor es el suavizado de Laplace, más irrelevantes son las probabilidades obtenidas de los datos y por lo tanto se obtiene menos precisión.

Por último se ha evaluado el mismo clasificador de distribución multinomial, pero esta vez sin suavizado de Laplace.

Se ha obtenido que el valor de `f1_score` es 0.986, por lo tanto inferior al logrado con un suavizado de 0.01.

```
size=0.0, f1_score=0.9861587864023917
```