

## Práctica 3: k-NN, PCA y Gauss

---

En esta práctica, se ha experimentado con los algoritmos vistos en las clases de teoría, de los clasificadores NearNeighbour (NN) y Gauss, además del algoritmo de reducción de dimensionalidad Principal Component Analysis (PCA).

Estos algoritmos, han sido utilizados sobre los conjuntos de datos dados:

- expressions
- gauss2D
- gender
- iris
- news
- ocr20x20
- videos

Para cada uno de estos conjuntos, se nos ofrecen dos subconjuntos, uno para la fase de entrenamiento, y otro para la fase de prueba.

Una vez aplicados los algoritmos, el error estimado por *hold-out* de estos, se muestran en la siguiente tabla:

### Tabla de errores de los clasificadores

TAREA	NN	PCA-NN	PCA-GAUSS
expressions	4.2	4.2	6.2
gauss2D	8.8	8.8	6.8
gender	5.1	5.3	48.1
iris	6.8	11.4	5.3
news	28.8	29.7	54.6
ocr20x20	0.7	0.7	0.3
videos	3.9	3.6	49.7

Como se puede observar en los valores de la tabla, la aplicación de PCA para el posterior uso del algoritmo de clasificación NN, aumenta el error estimado para todos los conjuntos de datos, exceptuando el caso de *“videos”*, en el cual, el error estimado decrementa respecto a la aplicación simplemente del algoritmo de clasificación.

Enfrentando ambos algoritmos de clasificación, para el caso de los volúmenes de datos más grandes, parece ser que el clasificador de Gauss sale perdiendo, dando mayor porcentaje de error para dichos conjuntos que utilizando el clasificador por *“vecino más próximo”*. En cambio, Gauss, mejora el porcentaje de error para los conjuntos de menor tamaño, en comparación al clasificador NN.