

Laboratorio de k-NN, PCA y Gaussianas

David Barbas Rebollo

Reconocimiento de Formas y Aprendizaje Computacional
Noviembre, 2020

1 Introducción

En esta práctica se utilizan 3 técnicas del aprendizaje automático. En primer lugar, vemos k-vecinos más cercanos, que clasifica una muestra según el voto de las k muestras más cercanas en el espacio de representación. En segundo lugar, se utiliza PCA, una técnica para reducir la dimensionalidad de los datos mediante una transformación lineal que retiene la máxima variación posible de los datos. Por último, se utiliza un clasificador de Gaussianas. En este caso se asume que los datos siguen una distribución gaussiana multivariada. Se calculan sus parámetros a partir de las muestras disponible. Posteriormente, se utiliza como estimador de máxima verosimilitud.

El objetivo de la práctica es mostrar los resultados de error de clasificación obtenidos al aplicar k-vecinos más cercanos, PCA con k-NN y PCA con Gaussianas sobre los datos proporcionado para la práctica.

2 Resultados

La siguiente tabla muestra, para cada conjunto, la estimación del error para distintos conjuntos utilizando distintas técnicas de clasificación.

Se puede observar en la siguiente tabla como **k-vecinos** obtiene el mejor error para los conjuntos *gender* y *news*. Con **PCA-NN** se obtiene el mínimo error para *videos*. Con **PCA-Gauss** para *gauss2D*, *iris* y *ocr20x20*. En cambio para *expressions* tanto **k-vecinos** como **PCA-NN** obtienen el mínimo error.

Table 1: Comparación de la estimación del error, mediante *hold-out*, para distintos conjuntos utilizando distintas técnicas de clasificación.

Tarea	Error (%)		
	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