

Minería de Datos (2018)

Práctico 1: Introducción al Aprendizaje Automático

- 1) En la aplicación 1 (análisis de dígitos manuscritos), los datos fueron proyectados a dos dimensiones a los fines de poder visualizarlos. Sin embargo, también podrían ser usados como representación para tareas de clasificación. Analice que sucede en este caso, si la dimensionalidad se reduce a 2, 4, 8, 16 y 32 dimensiones (puede utilizar otros clasificadores si lo cree conveniente). En caso de haber una mejora con estas dimensiones reducidas identifique en cuales dígitos se ha producido una mejora más significativa.
- 2) Explique cómo se calcula la precision, el recall y la medida F1. Ejemplifique con los valores obtenidos para Colin Powell, utilizando la matriz de confusión graficada en la aplicación 2 (reconocimiento de rostros). En esta misma aplicación, analice cuales serían los resultados si el PCA en lugar de obtener 150 componentes, trabaja con 50, 100 y 300 componentes.
- 3) Calcular manualmente la precision y el recall para la clase “sci.space” en la aplicación 3 (clasificación de textos). Ver luego, con el comando apropiado, si esos valores son correctos.
- 4) Ejecute el ejemplo de clasificación de textos con todas las categorías del data set. Informe cual es la categoría en que se obtiene mayor precision, recall y medida F1. Verifique si en Weka se obtienen resultados similares.