



CURSO: CMP 0575 - TÓPICOS 2 (DATA MINING)

COLEGIO: POLITÉCNICO

Semestre: Primer Semestre 2018/2019 NRC: 1068

Tarea 8: Ejercicio usando PCA

Problema:

1. Dado el conjunto de datos “**imágenes de chinas**” proporcionado en el D2L (link <https://www.dropbox.com/sh/0zcmvwjid2l8l9q/AAD0w25yPIZlUjveJfyF97VBa?dl=0>), se desea aplicar un PCA para la clasificación (reconocimiento) de nuevas instancias (caras de chinas). Para la realización de la tarea se exige:
 - Usar el **PCA**. De forma tal que se transforme el espacio original al de las componentes que capturen la mayor varianza de los datos. Los auto-vectores encontrados (seleccionados) constituyen una representación vectorial de la componente que explica la variación en las caras del conjunto de entrenamiento. Por tanto solo unos pocos auto-vectores combinados linealmente podrían ser suficientes para representar el conjunto de entrenamiento original (95% varianza acumulada).
 - Hacer una **investigación** de aplicación del PCA para reconocimiento de rostros. Pues la implementación del multiplicador de Lagrange para encontrar los auto-valores y auto-vectores de la matriz de covarianza S podría consumir tiempo (*Singular value decomposition*).
 - Sugerencia: Convertir las imágenes en vectores columnas para formar la matriz X .
 - Es obligatorio mostrar la trazabilidad del método durante la ejecución del programa:
 - i. Mostrar el vector media (convertir a imagen)
 - ii. Mostrar los auto-vectores seleccionados (convertidos a imágenes)
 - iii. Mostrar la combinación lineal de auto-vectores para formar los vectores de características.
 - iv. Mostrar la precisión de clasificación sobre el conjunto de test proporcionado por el profesor (sobre 10 imágenes de prueba). La clasificación debe ser basada en la minimización de la distancia euclidiana entre cada vector de prueba con respecto a los vectores característicos formados por combinaciones lineales de los auto-valores y auto-vectores.
 - v. La programación del algoritmo utilizado puede ser implementado en cualquier lenguaje.
 - Preparar una presentación científica para presentar al profesor.
 - Cargar al D2L los códigos implementados (archivo compactado que incluye el ejecutable ej: el .JAR de java) y la presentación dentro del plazo de entrega.