



PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE
ESCUELA DE INGENIERÍA
DEPARTAMENTO DE CIENCIA DE LA COMPUTACIÓN

Tarea 1

IIC2433 – Minería de Datos

Fecha de Entrega: Miércoles 11 de Septiembre hasta las 23:59 hrs

En esta tarea pondremos en práctica el análisis de las componentes principales y el discriminante de Fisher usando Python (<http://www.python.org/>). Para ello deben desarrollar los siguientes pasos:

- Bajar la base de datos Iris desde el UCI Machine Learning Repository (<http://archive.ics.uci.edu/ml/datasets/Iris>). Estos datos están descritos en 4 variables o atributos y la clase (3 posibles clases).
- Graficar los datos proyectados en dos dimensiones, usando todas las combinaciones posibles de dos variables (1-2 , 1-3 , 1-4 , 2-3 , 2-4 , 3-4), use colores para distinguir los datos de cada clase (Use la librería matplotlib de Python: <http://matplotlib.org/>).
- Implemente un código que calcule las componentes principales, ejecute su código en la base de datos Iris. Grafique los datos proyectados en la primera componente principal y en las dos primeras componentes principales. Nuevamente use colores para diferenciar entre las distintas clases. Comente los resultados.
- Implemente un código que calcule la mejor proyección de los datos usando el discriminante de Fisher. Para esto asuma 3 problemas de dos clases: i) clase 1 ó cualquier otra ii) clase 2 ó cualquier otra iii) clase 3 ó cualquier otra.

En cada problema ejecute el código para calcular el discriminante de Fisher y obtenga el vector donde debe proyectar los datos. Haga la proyección y grafique, muestre en colores cada una de las **dos** clases (la clase “i” o la clase “cualquier otra”). No olvide que esto se realiza tres veces, una para cada clase. Comente los resultados.

La tarea debe ser entregada en un archivo comprimido que contenga: el código Python, un .pdf con el informe y un archivo Readme.txt con las instrucciones para ejecutar el código. El código debe estar debidamente comentado. Se habilitará un buzón en el sitio del curso cerca de la fecha de entrega para que puedan subir su archivo, no se aceptan archivos atrasados, la tarea es **individual**.