

# Tarea n°4

1. Retomamos el conjunto de datos de Iris que podemos llamar gracias a la función `load_iris` del paquete `sklearn.datasets`. En estudios de datos anteriores vimos que las dos variables `petal length` y `petal width` son muy correladas con las categorías de las plantas.
  - (a) Usando los datos bidimensionales correspondientes a `petal length` y `petal width` de cada planta, observar el resultado del clustering  $k$ -medias de los datos con 2, 3 y 4 clases. Se usará la función `KMeans` de `sklearn.cluster`.
  - (b) Escribir una función `Riesgo(X,y,C)` que toma los datos  $X$ , las categorías correspondientes  $y$  y los centroides  $C$  y que responde el valor de la función  $G$  (definida en el curso).
  - (c) Graficar el resultado de `Riesgo` para  $k$  que varia entre 2 y 10.
  - (d) Para elegir el mejor  $k$  una técnica famosa es la Elbow Method. Hacer una búsqueda de referencias que explican la técnica. Implementar la técnica para la grafica del inciso anterior.