



La técnica de vecinos más cercanos es uno de los métodos de minería de datos más simple que se puede imaginar. Con este método, se puede predecir información (que no disponemos) asociada a un determinado registro de datos, simplemente mediante las informaciones (que sí tenemos) de los registros de datos que más se parezcan a él.

La técnica de vecinos es un método general que se puede aplicar a cualquier conjunto de datos. En este ejercicio trabajaremos sobre el dataset IRIS. Este [dataset](#) contiene 150 registros de medidas de flores (lirios) y para cada uno de estos registros se proporciona la información adicional de la especie a la que pertenece. Hay tres especies distintas: setosa, versicolor y virginica.



Iris setosa



Iris versicolor



Iris virginica

Los datos de entrada estarán en formato CSV, estas son las primeras líneas del fichero de entrada:

```
sepal_length,sepal_width,petal_length,petal_width,species
5.1,3.5,1.4,0.2,setosa
4.9,3,1.4,0.2,setosa
4.7,3.2,1.3,0.2,setosa
4.6,3.1,1.5,0.2,setosa
```

El dataset completo se encuentra en el fichero IRIS.csv de la carpeta /datos. Además, hay dos ficheros IRIS_test.csv e IRIS_train.csv que contienen los mismos registros que IRIS.csv distribuidos en dos grupos de 60 y 90 registros, respectivamente.

En este ejercicio implementaremos un clasificador basado en la técnica de vecinos más cercanos con dos posibles estrategias: votación y ponderación. Implementaremos también funciones que permitirán evaluar la calidad del clasificador obtenido, y mostraremos gráficamente los vecinos más cercanos a un determinado registro del dataset IRIS.

