

Vision Artificial. GIEC.

Sistemas de Vision Artificial. GIC.

Miguel Angel Garcia, Juan Manuel Miguel, Sira Palazuelos.

Departamento de Electrónica. Universidad de Alcalá.

Tema 5: ejercicio 03 - Kmeans

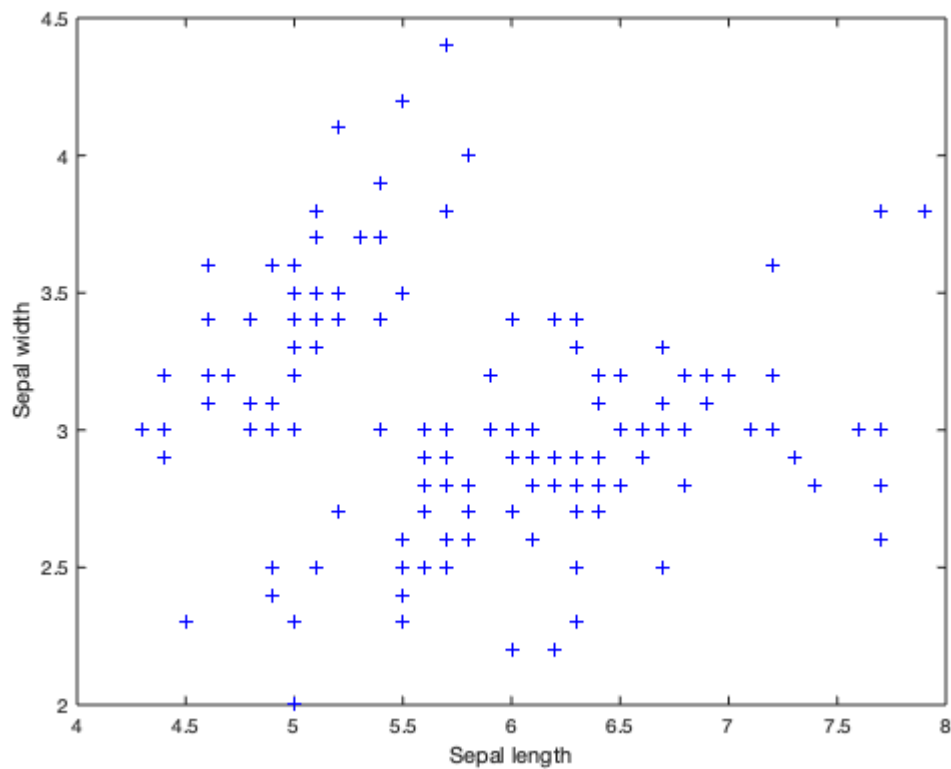
La base de datos **fisheriris** consiste en mediciones de (1) la longitud del sépalo, (2) la anchura del sépalo, (3) la longitud del pétalo y (4) la anchura del pétalo para 150 especímenes de iris (el iris es un conjunto de especies de plantas con flores llamativas). Hay 50 especímenes de cada una de las tres especies {**setosa**, **versicolor** y **virginica**}.

Step 1) Carga datos

```
clear all;  
close all;  
  
load fisheriris  
data_iris = meas(:,1:2);  
label_iris = categorical(species);  
names_iris = categories(label_iris);
```

Step 2) Visualiza datos.

```
plot(data_iris(:,1),data_iris(:,2),'b+','MarkerSize',5);  
xlabel('Sepal length');  
ylabel('Sepal width');
```



Step 3) Kmeans

```
k = 3;
[idx,C] = kmeans(data_iris,k);
```

Step 4) Plot clustering.

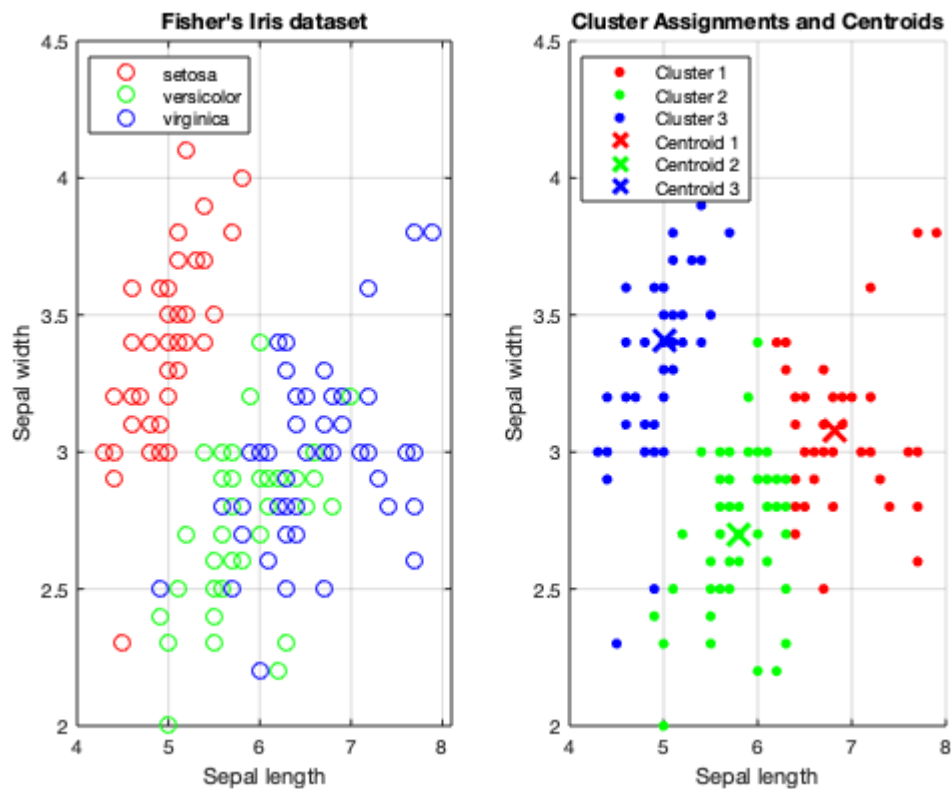
```
figure;
subplot(1,2,1)
gscatter(data_iris(:,1),data_iris(:,2), label_iris,'rgb','ooo',8,'on')
xlabel('Sepal length');
ylabel('Sepal width');
grid on;
legend('setosa','versicolor','virginica','Location','NW')
title('Fisher's Iris dataset')

subplot(1,2,2)
hold on
plot(data_iris(idx==1,1),data_iris(idx==1,2),'r.','MarkerSize',12)
plot(data_iris(idx==2,1),data_iris(idx==2,2),'g.','MarkerSize',12)
plot(data_iris(idx==3,1),data_iris(idx==3,2),'b.','MarkerSize',12)
plot(C(1,1),C(1,2),'rx','MarkerSize',15,'LineWidth',3)
plot(C(2,1),C(2,2),'gx','MarkerSize',15,'LineWidth',3)
plot(C(3,1),C(3,2),'bx','MarkerSize',15,'LineWidth',3)
xlabel('Sepal length');
```

```

ylabel('Sepal width');
grid on;
legend('Cluster 1','Cluster 2','Cluster 3','Centroid 1','Centroid 2','Centroid 3','Location');
title('Cluster Assignments and Centroids')
hold off

```



Se pide:

1. Calcule la matriz de confusión del algoritmo K-means usando la función `confusionmat`. Tenga en cuenta que este no es un algoritmo supervisado por lo que las etiquetas pueden estar cambiadas ¿cómo influye esto en la matriz de confusión?
2. Calcular la precisión ($\text{acc} = \text{hits}/\text{total}$) del algoritmo K-means.
3. Repita las secciones anteriores pero usando pétalos en lugar de sépalos (`data_iris = meas(:,3:4);`). ¿En qué caso se obtienen mejores resultados?