



Universidad Simón Bolívar
Departamento de Computación

Inteligencia Artificial III (CI5437)
Proyecto 3: Clustering.

Integrantes:
Isaac González 11-10396
Andrea Centeno 10-10138

Sartenejas, marzo 2017.

ACTIVIDAD 2.a

Luego de explorar los datos de *Iris Data Set* con el algoritmo de K-means con $k=2,3,4,5$, notamos que en varios casos las etiquetas coinciden con las reales, sobretodo cuando se trata del conjunto *Iris-setosa*. Los puntos de dicho conjunto suele ser identificados sin error luego de correr K-means con $k > 2$. Sin embargo, en todas las corridas se nota que en, al menos un cluster, el algoritmo confunde los valores de las etiquetas Iris-virginica y Iris-versicolor.

Para $K = 2$ tenemos que el algoritmo logra diferenciar dos conjuntos: un conjunto que contienen aquellos puntos con etiqueta Iris-setosa y otro con aquellos con etiquetas Iris-virginica y Iris-versicolor. A diferencia con las etiquetas reales, el algoritmo falla 3 de 150 puntos. El cluster que contiene la etiqueta Iris-setosa, incluye 3 puntos con etiqueta Iris-versicolor que deberían pertenecer al otro grupo.

	Iris-setosa	Iris-versicolor	Iris-virginica	Número de puntos reconocidos por el algoritmo	Número de puntos reales
Cluster 1	50	3		53	50
Cluster 2		47	50	97	100

El número en rojo, representa la cantidad de puntos que no deberían estar en el cluster 1.

Para $k = 3$ tenemos que el algoritmo reconoce 3 grupos de datos, uno es el grupo que contiene a Iris-setosa, el cual reconoce perfectamente, los otros dos grupos no son reconocidos por completo, uno de los grupos restantes debería contener sólo puntos con etiqueta Iris-virginica y el otro, puntos con etiquetas Iris-versicolor.

	Iris-setosa	Iris-versicolor	Iris-virginica	Número de puntos reconocidos por el algoritmo	Número de puntos reales
Cluster 1		2	36	38	50
Cluster 2	50			50	50
Cluster 3		48	14	62	50

Los números en rojo representan la cantidad de puntos que no deberían estar en el cluster dado, ya que pertenecen a otro grupo de datos.

Los resultados de este ejercicio se encuentran en el archivo *resultados_act2a.txt*

ACTIVIDAD 2.b

Para la actividad 2.b realizamos una pequeña modificación al algoritmo k-means desarrollado en la actividad 1 para que trabajara con píxeles en lugar de puntos, esto se debe a que queríamos almacenar las coordenadas del píxel y sus colores en RGB. El algoritmo *kmeans_image* utiliza el color de los píxeles para agruparlos en k-clusters, para realizar la comprensión de la imagen, una vez divididos los píxeles en k-clusters, procedemos a pintar los píxeles de un cluster del color que indica el centroide obteniendo así una imagen de k colores.