



# Tecnológico de Monterrey

ITESM CSF

Aprendizaje automático 7-10 lu

Dr. Víctor de la Cueva

Noviembre 9 2017

Adrián Biller A01018940

**Documentación Proyecto 7**

## Introducción

Este proyecto fue realizado con base en el material obtenido sobre el aprendizaje no supervisado, más específicamente sobre el algoritmo de kmeans con el cual se puede separar diferentes datos en clusters. Este programa lee un archivo con 300 datos y separa dichos datos en clusters.

## Manual de usuario

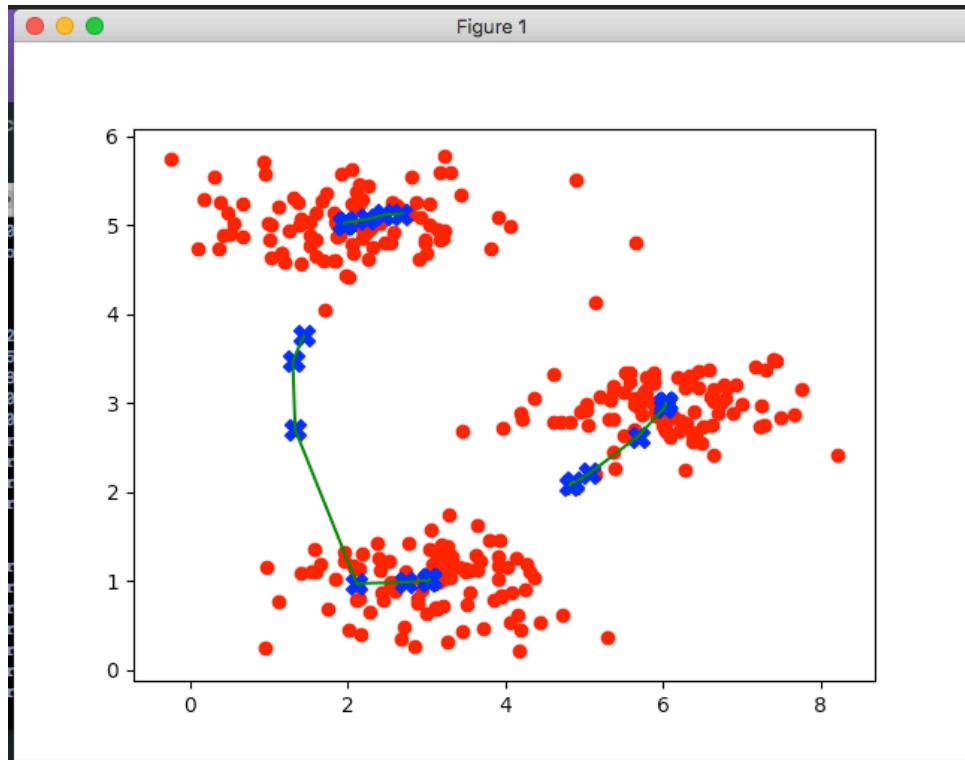
Para poder utilizar este programa se debe utilizar la terminal. Antes la computadora en donde se realice la prueba debe contener Python 3 (el programa fue realizado utilizando Python 3.6.1) además se utilizan las librerías *random*, *numpy*, *matplotlib.pyplot* y *pylab* en el programa, en caso de errores con las librerías se deberán instalar por separado dependiendo de la máquina que se esté utilizando puede variar el método de instalación.

Primero se deberá colocar en la carpeta en la que se localice el programa .py y el archivo .txt conteniendo los datos a analizar llamado "ex7data2.txt". Una vez dentro de la carpeta se deberá ingresar el siguiente comando en terminal:

```
$ python3 proyecto7.py
```

Una vez ingresado el comando se correrá el programa y regresará el resultado y se mostrarán en pantalla, en caso de querer repetir el proceso se deberá realizar el mismo comando en terminal.

Al correr el programa se mostrará una gráfica con los cambios de centroides por cada iteración. De esta manera se pueden ver los diferentes clusters y como se van cambiando los centroides al centro de cada cluster.



En el main de este programa se encuentran solamente la llamada a la función `kMeansInitCentroids` que recibe la matriz `x` que son los valores y el número de clusters, es decir `k`.

Posteriormente se llama a la función `runKMeans` donde se envía la matriz con los datos leídos `x`, los centroides iniciales, el número de iteraciones que estará corriendo el programa y un valor `true` o `false` dependiendo si se desea graficar los datos como mostrado anteriormente.

## Funciones

Además de las funciones ya especificadas en el documento se creó una función llamada `plot_clusters` que fue basada de un ejemplo en internet en la cual se puede graficar todos los puntos y el camino que van tomando los centroides.