



Fundamentos de Aprendizaje de Máquina

PhD Jorge Rudas



K-means

```
1. Randomly initialize the k cluster centroids  
repeat{  
  2. Assign the data points to cluster centroids  
  3. Move cluster centroids  
}
```

Examples/kmeans.ipynb

Pruebe con datos sintéticos los siguientes experimentos

1. **(0.5 punto)** 1000 muestras, 2D, $k = 3$, distancia euclidiana, max iteraciones = 10, 100, 1000, 10000
2. **(0.5 punto)** 1000 muestras, 3D, $k = 3$, distancia euclidiana, max iteraciones = 10, 100, 1000, 10000
3. **(0.5 punto)** 1000 muestras, 10D, $k = 3$, distancia euclidiana, max iteraciones = 10, 100, 1000, 10000
4. **(0.5 punto)** 1000 muestras, 100D, $k = 3$, distancia euclidiana, max iteraciones = 10, 100, 1000, 10000
5. **(0.5 punto)** (0.5) 1000 muestras, 3D, $k = 2, 3, 5, 10$ distancia euclidiana, max iteraciones = 1000
6. **(1 punto)** 1000 muestras, 2D, $k = 5$, distancia euclidiana, manhattan y mahalanobis, max iteraciones = 1000
7. **(2 punto)** Determine para cada experimento, cuántas iteraciones son necesarias para la convergencia (centroides actuales son iguales a centroides anteriores)
8. **(1 punto)** Presentación

Realice al menos 5 corridas para cada experimento y reporte el resultado medio. Discuta con su equipo sus resultados, saque sus conclusiones, realice una presentación de 2 slices y socialice con su clase.

90 min de taller + 45 presentaciones

Actividad opcional

https://github.com/dkarunakaran/kmeans_clustering/blob/master/kmeans.ipynb

Pruebe con una imagen de alguna locación en Ciénaga

- Proponga una estrategia para determinar el mejor valor de k
- Se pueden mejorar los resultados si se hace una selección NO aleatoria de los centroides? Qué propondría?