

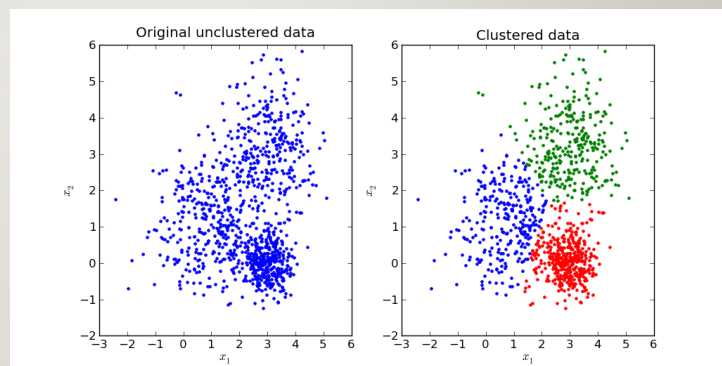
K-MEANS

ALGORITMO DE CLUSTERING

1

K-MEANS

- Algoritmo de aprendizaje no supervisado. Tratamos de hacer sentido de la estructura de datos sin etiquetas.
- Agrupa vectores de acuerdo con su cercanía a una cantidad K de centroides que conforman buenas representaciones de los datos.



Fuente: <https://i.stack.imgur.com/cIDB3.png>

2

APLICACIONES DE K-MEANS

Segmentación
de Mercado

Categorización
de inventario

Agrupando
mediciones de
sensores

Detección de
anomalías

Compresión
de datos

3

PSEUDOCÓDIGO

```

Initialize  $\mathbf{m}_i, i = 1, \dots, k$ , for example, to  $k$  random  $\mathbf{x}^t$ 
Repeat
  For all  $\mathbf{x}^t \in \mathcal{X}$ 
     $b_i^t \leftarrow \begin{cases} 1 & \text{if } \|\mathbf{x}^t - \mathbf{m}_i\| = \min_j \|\mathbf{x}^t - \mathbf{m}_j\| \\ 0 & \text{otherwise} \end{cases}$ 
  For all  $\mathbf{m}_i, i = 1, \dots, k$ 
     $\mathbf{m}_i \leftarrow \sum_t b_i^t \mathbf{x}^t / \sum_t b_i^t$ 
Until  $\mathbf{m}_i$  converge
  
```

Figure 7.3 k-means algorithm.

$$B = \begin{bmatrix} b_1^1 & b_1^2 & \dots & b_1^K \\ b_2^1 & b_2^2 & \dots & b_2^K \\ \vdots & \vdots & \ddots & \vdots \\ b_N^1 & b_N^2 & \dots & b_N^K \end{bmatrix}$$

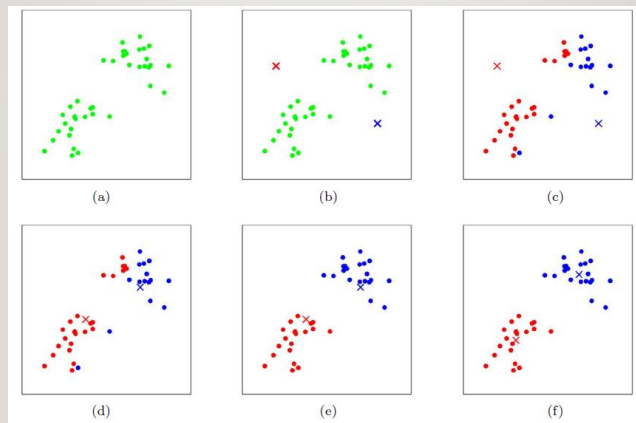
$$B = \begin{bmatrix} 0 & 1 \\ 0 & 1 \\ 1 & 0 \end{bmatrix}$$

Notas: $i = \{1, \dots, K\}$, donde K representa el número de clusters. \mathbf{m}_i representa los centroides (medias) de los clusters, $\mathbf{x}^t \in \mathcal{X}$ para $t = \{1, \dots, |\mathcal{X}|\}$ son los vectores de datos y b_i^t indica las asociaciones de cada vector de datos a un cluster.

Fuente: Alpaydin, E. "Introduction to Machine Learning", 2 ed.)

4

ILUSTRACIÓN DE UNA ITERACIÓN DEL ALGORITMO PARA K=2



Fuente: Ng, A. Stanford CS229: Lecture Notes

5

NOTAS PARA IMPLEMENTACIÓN

- K-means requiere que el usuario especifique el valor de K.
- Trate de reducir los movimientos de datos en la memoria y la cantidad de bucles.
- Criterios de convergencia para implementar K-means:
 - Cantidad de iteraciones.
 - Error menor que un umbral.
 - Actualización de un parámetro es menor que un umbral.
 - Cantidad de memoria utilizada.
- K-means tiene un comportamiento determinístico, y se recomienda utilizar varias corridas empleando inicializaciones aleatorias para obtener resultados concluyentes.

6

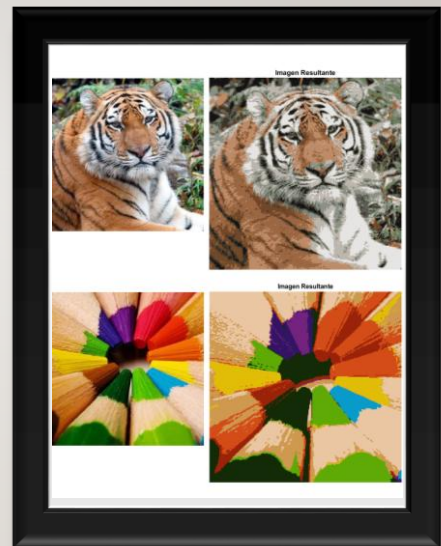
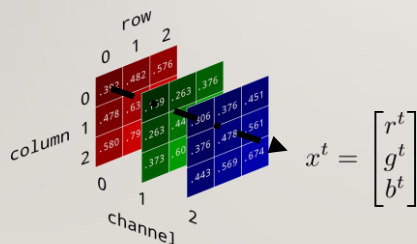
NOTAS PARA IMPLEMENTACIÓN

- K-means require que el usuario especifique el valor de K.
- Trate de reducir los movimientos de datos en la memoria y la cantidad de bucles.
- Criterios de convergencia para implementar K-means:
 - Cantidad de iteraciones.
 - Error menor que un umbral.
 - Actualización de un parámetro es menor que un umbral.
 - Cantidad de memoria utilizada.
- K-means tiene un comportamiento determinístico, y se recomienda utilizar varias corridas empleando inicializaciones aleatorias para obtener resultados concluyentes.

7

APLICACIÓN A LA COMPRESIÓN DE IMÁGENES

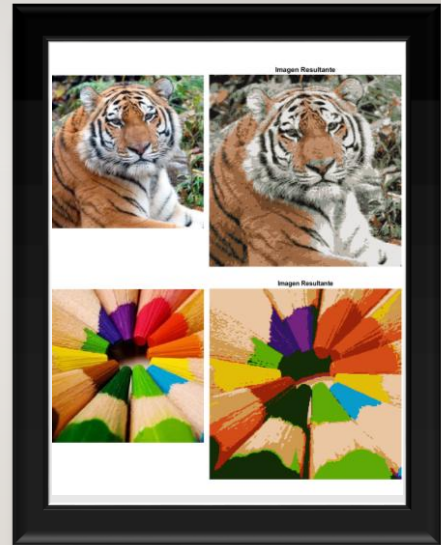
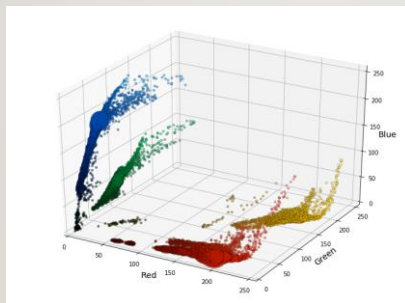
Construcción de vectores de features:



8

APLICACIÓN A LA COMPRESIÓN DE IMÁGENES

Visualización:



Fuente: <https://tberg.dk/post/determining-dominant-colors/>