

Clustering Algorithms comparision

	K-means	K-medoids	DBSCAN	Spectral clustering
Principio matemático	Optimización de la función objetivo (mínimos cuadrados)	Optimización de la función objetivo (distancia manhattan o Euclidiana)	Densidad de puntos	Descomposición Espectral
Descripción	Divide los datos en K clusters minimizando la varianza dentro de cada uno.	Divide los datos en K clusters minimizando la varianza dentro de cada uno.	<u>Utiliza</u> la densidad de los puntos vecinos para identificar clusters conectados.	Utiliza la descomposición espectral de la matriz de afinidad para identificar clusters
Desventajas	<ul style="list-style-type: none"> - No funciona bien con clusters de tamaños y formas irregulares. - Sensible a la inicialización de los centroides y puede converger a un mínimo local. 	<ul style="list-style-type: none"> - más lento que k-means por el cálculo de las <u>distancias</u> entre medoides. 	<ul style="list-style-type: none"> - Sensible a los parámetros eps y min_samples - No funciona bien con datos de alta dimensionalidad 	<ul style="list-style-type: none"> - Sensible a la elección del número de clusters y la construcción de la matriz de similitud.
Ventajas	<ul style="list-style-type: none"> - Escalable a grandes dimensiones - Eficiente en grandes conjuntos de datos 	<ul style="list-style-type: none"> - Adecuado para datos con presencia de outliers - Menos sensible a la inicialización que KMeans debido a la selección de medoide 	<ul style="list-style-type: none"> - No hay que definir el número de clusters previamente - Detecta a outliers - Clúster en formas que no sean esféricas 	<ul style="list-style-type: none"> - Robusto a la inicialización de los centroides.
Aplicaciones	<ul style="list-style-type: none"> - Segmentación de clientes - Reconocimiento de patrones 	<ul style="list-style-type: none"> - Análisis de redes sociales 	<ul style="list-style-type: none"> - Imágenes - Datos geográficos - Agrupas usuarios en RRSS 	<ul style="list-style-type: none"> - Clasificación de imágenes - Reconocimiento de patrones
Computacionalmente	Eficiente	Eficiente	Moderadamente costoso	Costoso debido al cálculo de la matriz de similitud y la descomposición espectral.

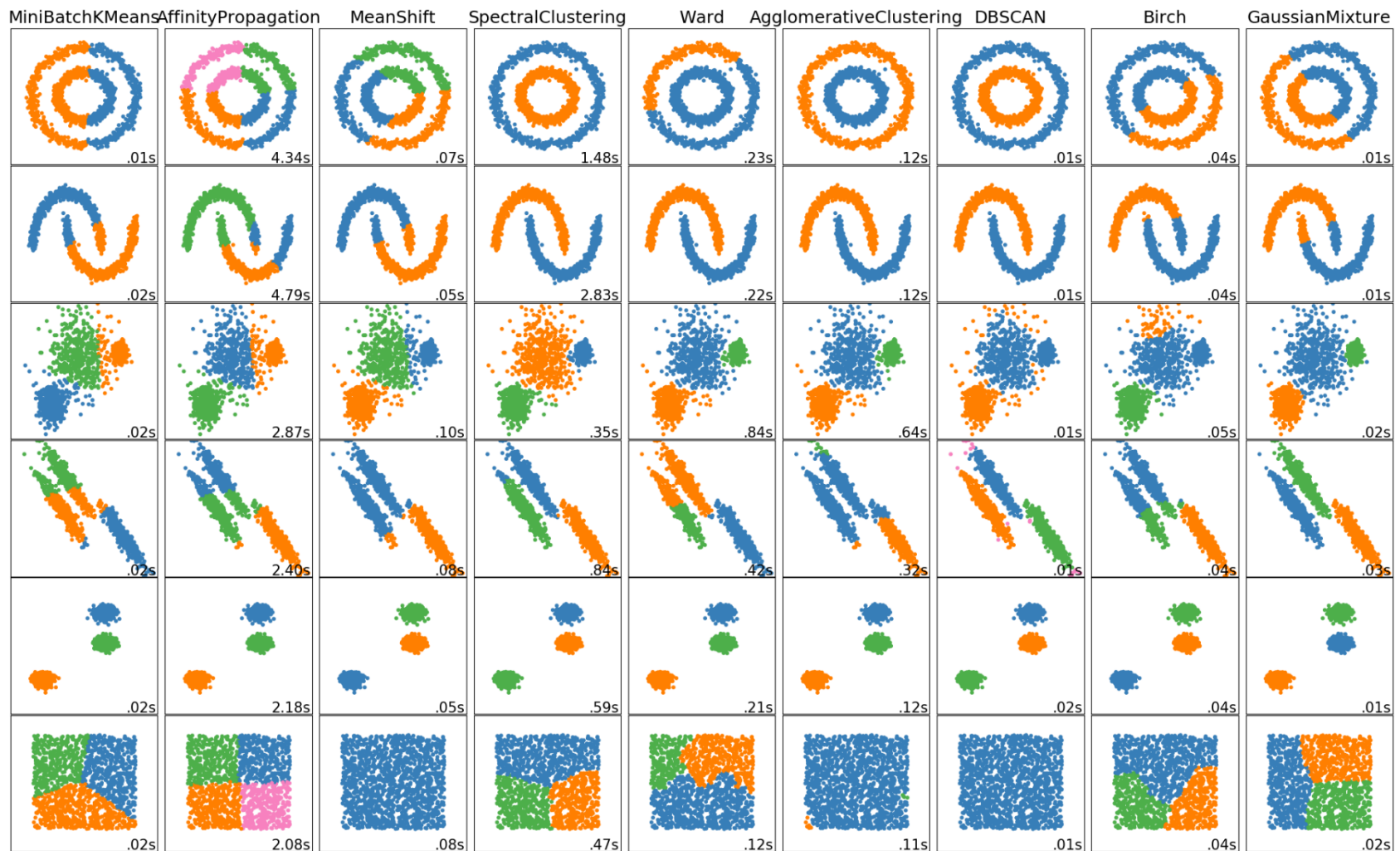


Imagen tomada de: <https://towardsdatascience.com/the-5-clustering-algorithms-data-scientists-need-to-know-a36d136ef68>