

Tarea 03 - Reconocimiento de patrones
Giovanni Gamaliel López Padilla

La base de datos **Animales con Atributos** (*Animals with Attributes*) contiene información sobre 50 animales. Para cada uno, se tienen 85 características de valor real que capturan varias propiedades del animal: dónde vive, qué come, etc. Esta es claramente una aplicación en la que esperamos que la agrupación descubra “grupos naturales” en los datos.

- Usa ISOMAP, LLE, T-SNE y SOM's para encontrar visualizaciones informativas de los datos.
- Aplique el algoritmo de Clustering Jerárquico usando las funciones linkage: single, complete, ward.

Muestre los dendogramas correspondientes y responda las siguientes preguntas basado en los resultados obtenidos

1. Hasta qué punto tienen éxito estos algoritmos en esta tarea?
2. Son los grupos en su mayoría razonables?
3. Podemos, en general, esperar que la agrupación capte perfectamente lo que queremos?
4. Para cual(es) función(es) de linkage estaría satisfecho con el agrupamiento? Comente su respuesta.

El resultado del análisis debe ser un pequeño reporte con textos e imágenes integrados.

Visualizaciones informativas

A partir de los datos contenidos en el archivo [predicate-matrix-continuous.txt](#) se realizaron los procesos de Isomap, LLE, T-SNE y SOM para obtener visualizaciones de la distribución de los datos.

Isomap

En la figura 1 se visualizan los resultados que obtuvo al aplicar isomap. En estos resultados se visualiza que existe una diferencia entre ciertos grupos de animales. Por ejemplo en la parte superior derecha se realizó la agrupación del gorila y el chimpancé. En la parte inferior derecha animales como la morca y la orca también se encuentran cercanos. En la parte central se puede ver que realizó la agrupación de animales domésticos junto a animales que pertenecen a la misma familia.



Figura 1: Resultados de aplica el isomap a los datos contenidos en el archivo [predicate-matrix-continuous.txt](#).

LLE

En la figura 2 se visualiza el resultado que se obtuvo al aplicar LLE a los datos. A diferencia del método de isomap, LLE realizo una mayor distinción entre animales marítimos y terrestres. Realizando un grupo de animales pertenecientes a las granjas y primates. En otros casos no llegó a realizar la distinción entre animales domésticos.



Figura 2: Resultados de aplica LLE a los datos contenidos en el archivo [predicate-matrix-continuous.txt](#).

T-SNE

En la figura 3 se observan los resultados que se obtuvo al aplicar T-SNE a los datos. A diferencia de los métodos anteriores, en T-SNE no se logra visualizar agrupamientos de animales con gran distinción. Se observa que los animales acuaticos ocupan la parte de la derecha. Los primates se encuentran en la parte superior y animales salvajes como el tigre y león se encuentran en la parte de la izquierda. Revisando más a detalle se observa que la mayoría de los animales domésticos se encuentran en la parte central.

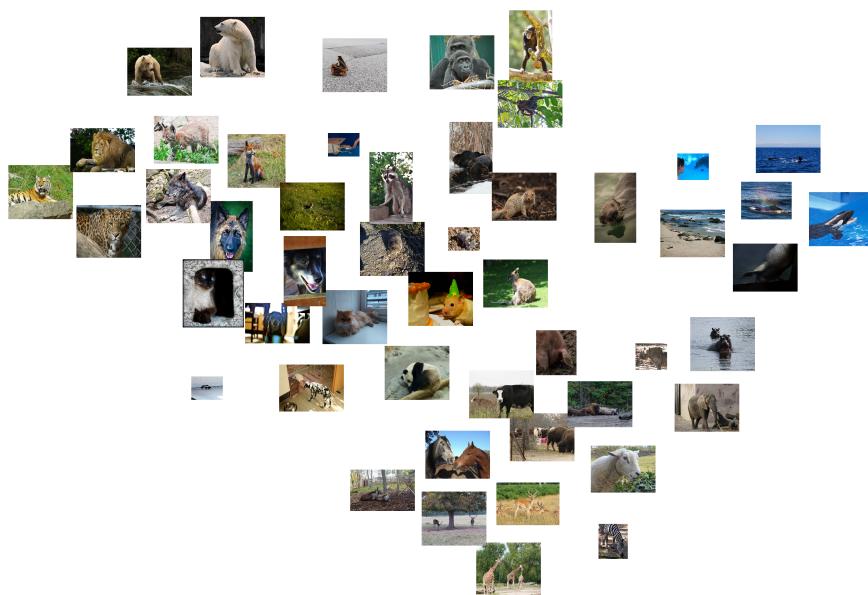


Figura 3: Resultados de aplica LLE a los datos contenidos en el archivo [predicate-matrix-continuous.txt](#).

SOM

Los métodos anteriores realizan una reducción de dimensionalidad en los datos. Esto funciona para reconocer si existen grupos en los datos. En la figura 4 se representan los grupos encontrados por SOM en la parte izquierda y en la parte derecha la posición de cada animal en cada reducción de orden antes calculada. Con SOM se logra delimitar de una mejor manera los resultados obtenidos en cada reducción de orden identificando a cada animal en una clase.

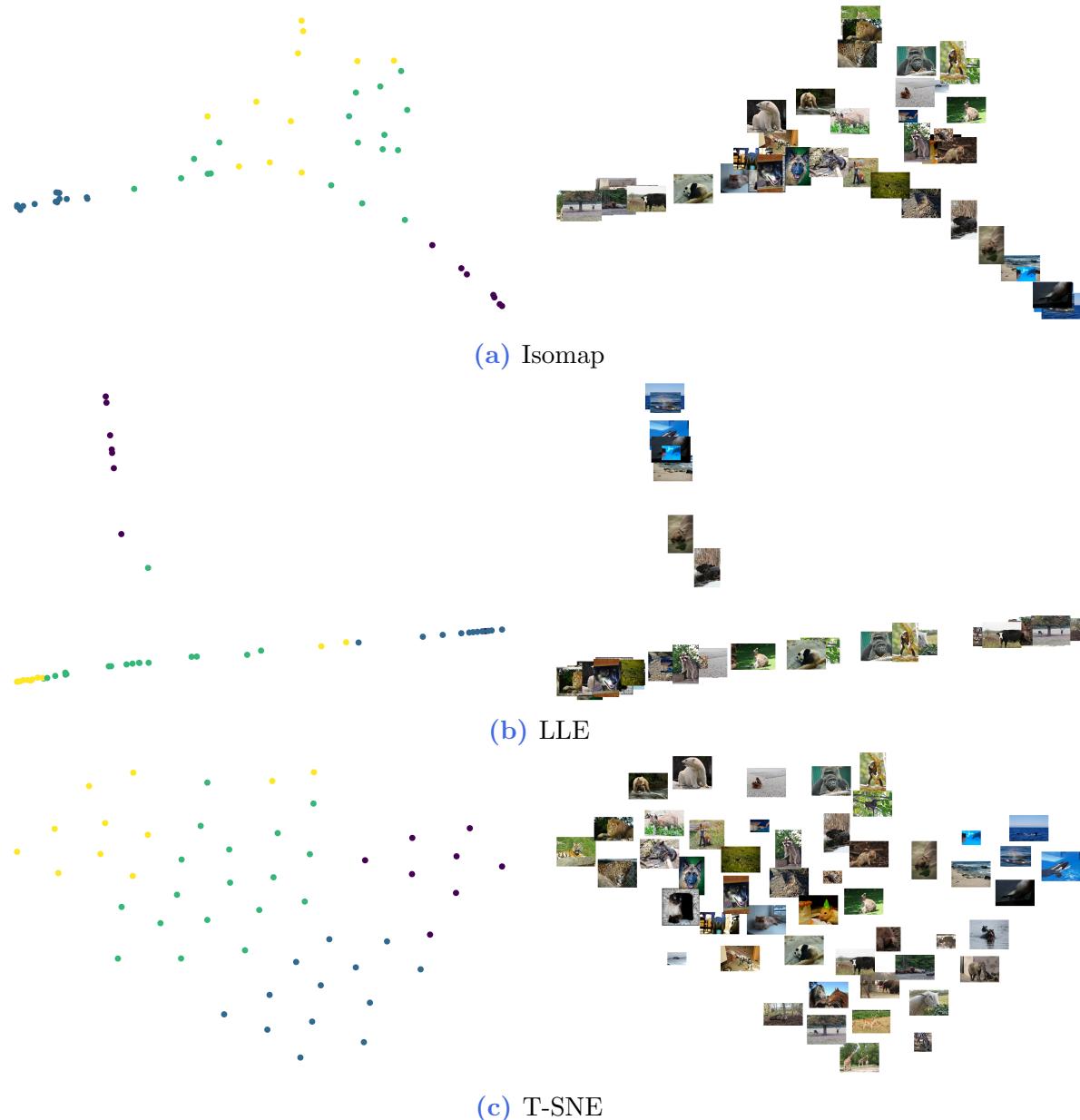
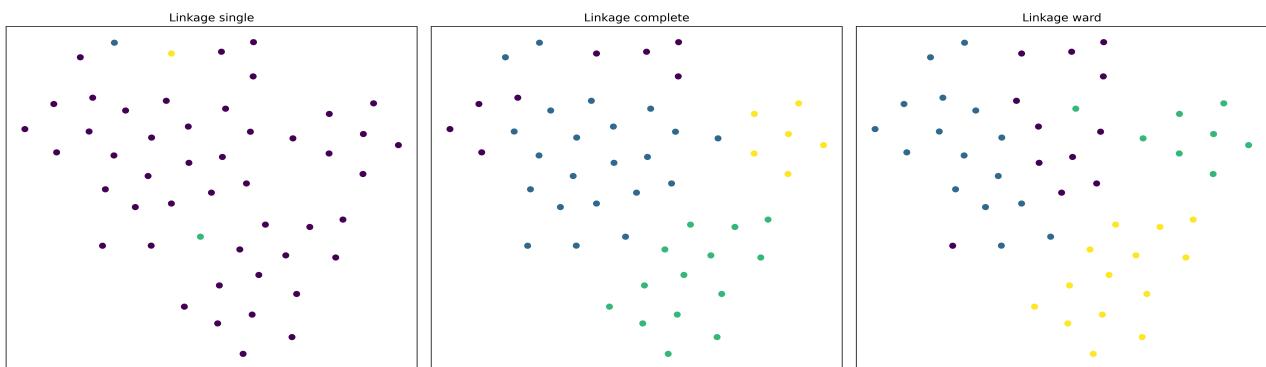


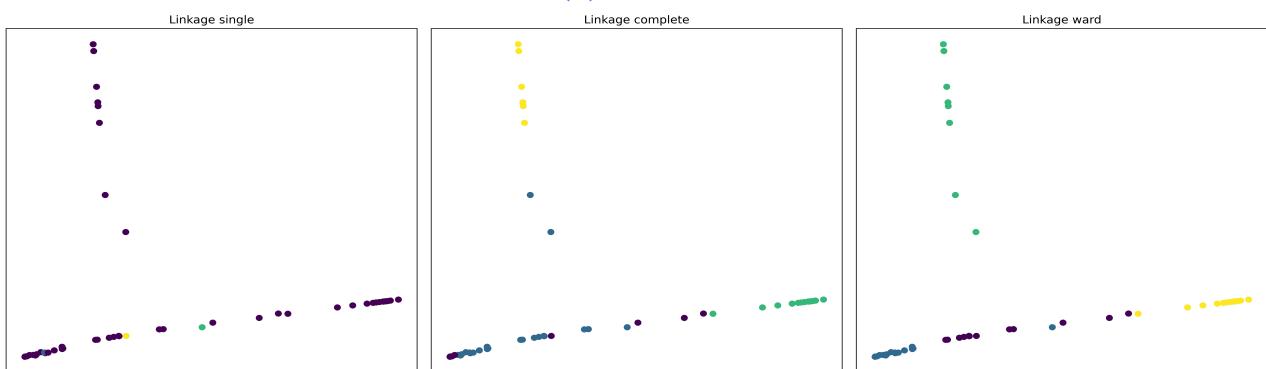
Figura 4: Resultados de aplicar SOM a el conjunto de datos con cada reducción de dimensión

Clustering

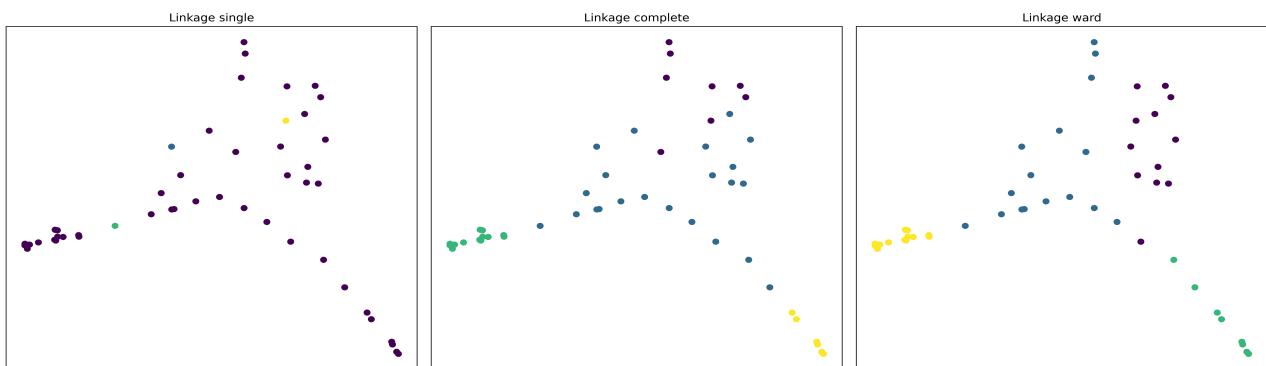
Con los resultados de SOM es razonable aplicar más algoritmos para asignar clases a cada uno de los animales para conseguir una mejor delimitación en cada reducción de orden. En la figura 5 se visualizan los resultados de aplicar clustering jerárquico de los tipos tipos single, complete y ward con las reducciones de dimension isomap, LLE y T-SNE.



(a) Isomap.



(b) LLE.



(c) T-SNE.

Figura 5: Clustering jerárquico aplicado en conjunto a la reducción de dimensión.

Con esto se visualiza que los tipos complete y ward obtienen resultados semejantes y con buenas representaciones al hacer una comparación de estos valores con las imágenes de cada animal como se muestran en las figuras 1, 2 y 3.

En la figura 6 se visualizan los денограммas de cada clustering jerárquico aplicado a cada tipo.

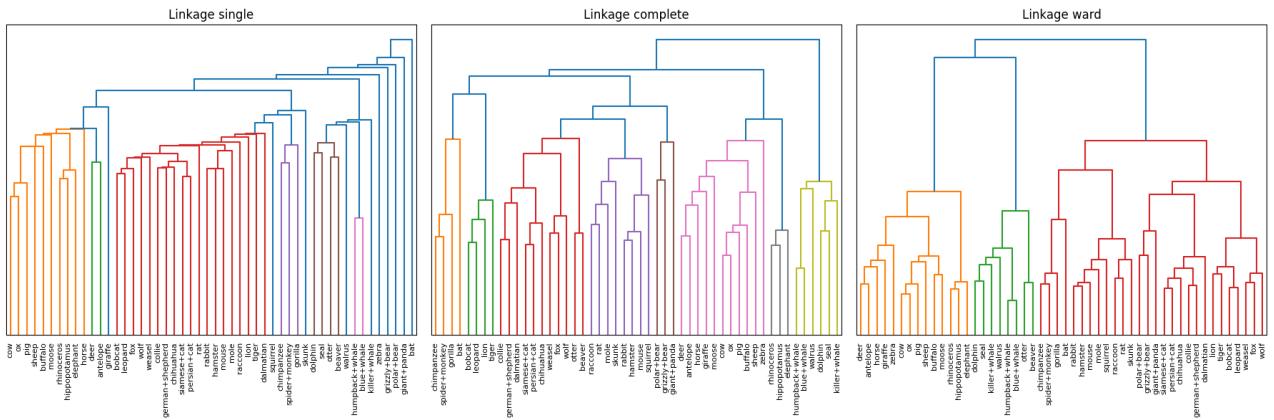


Figura 6: Denogramas de los clustering jerarquico de tipo single, complete y ward.

Con estos resultados se obtiene que los modelos de clustering jerárquico del tipo complete y ward realizan un buen trabajo realizando la asignación de categorías a cada elemento que tenemos. Donde el tipo ward tiene una representación en денограма más sencilla de visualizar e interpretar. Estos modelos llegan a realizar una categorización pero la interpretación que tiene se la damos nosotros, por lo que no se puede especificar sobre qué categoría o atributos se pueda hacer un énfasis.