

Elements of Machine Learning 2025

Lista 03

30.abril.2025

1. ¿Cuál método es más sensible a datos atípicos: k -medias o agrupamiento jerárquico? Motivar ampliamente su respuesta con ejemplos.

¿Cómo se puede resolver el problema de los datos atípicos a la hora de realizar un método de agrupamiento. Exponga sus ideas y evidéncielas a través de un ejemplo.

2. Hacer un agrupamiento jerárquico con los datos de los países `countries_binary.xlsx`.

Visualizar los resultados o dendrogramas de diferentes métodos de agrupamiento, variando los siguientes:

- el método de agrupamiento (simple, completo, promedio, Ward)
- la métrica utilizada (euclídeana, Hamming).

3. Realizar un análisis de agrupamiento, nuevamente para los datos de los países, que están disponibles en el archivo `countries_binary.xlsx`. Utilice el o los algoritmos que usted considere convenientes.

Con estos datos, hacer la proyección PCA de los datos a 2 componentes principales, e ilustrar cómo se ven las regiones de clasificación obtenidas con su método de agrupamiento en este espacio 2-dimensional.

Ilustrar cómo cambian las regiones de clasificación cuando se utilizan diferentes métricas en el k -means.

4. Para al menos 2 imágenes RGB de su elección, realizar un algoritmo de cuantización de colores para la imagen, usando como base un algoritmo de k -medias, y un segundo algoritmo de clustering de su elección.

Ilustrar los resultados obtenidos comparando qué ocurre:

- al aumentar el número k de grupos.
- al modificar la métrica o los parámetros usados en los algoritmos.

Compare sus resultados y elabore conclusiones.
