

DBSCAN

Evelyn Yosiana / 13522083

1. Konsep DBSCAN intinya yaitu pilih titik-titik yang belum masuk ke cluster manapun, cari tetangganya, dari tetangganya cari tetangganya lagi, begitu seterusnya sampai seluruh titik data sudah memiliki cluster.

Proses train: untuk proses train, iterasi tiap titik, cek apakah titik tersebut sudah masuk ke cluster tertentu (labelnya bukan -1), jika belum, lakukan:

- Cari titik data lain yang jaraknya lebih kecil daripada ϵ
- Kelompokkan titik data tersebut, hitung jumlah anggotanya, jika jumlah anggotanya $> \text{min_samples}$, ubah label data-data tersebut menjadi satu cluster. Jika sebaliknya, ubah label data di cluster tersebut menjadi -1 lagi.

Parameter:

- ϵ : jarak maksimal agar dapat dianggap tetangga (secluster).
 - min_samples : minimal anggota dalam satu cluster.
 - metric : metrik perhitungan jarak, ada 3 (euclidean, manhattan, minkowski).
 - p : parameter order untuk minkowski.
4. Pada model ini, hasil perbandingan model yang saya buat dan model dari library kurang sama. Hal ini dapat disebabkan oleh beberapa faktor, antara lain:
 - **Cara menangani noise:** algoritma dalam library memiliki aturan tertentu untuk noise (apakah tetap menjadi noise atau menjadi bagian dari suatu cluster), sedangkan dalam kode saya, noise yang tidak tersentuh oleh titik lainnya akan tetap dibiarkan menjadi noise.
 - **Urutan pemilihan titik:** dalam kode saya, pemilihan titik dilakukan secara terurut, namun kode dalam library mungkin menggunakan random state ataupun urutan titik yang berbeda.
 - **Optimasi:** kode dari library sudah dioptimasi, misalnya dengan menggunakan KD-Tree maupun Ball-Tree sehingga proses komputasinya lebih cepat dan mungkin saja kurang akurat untuk kasus tertentu.

Lampiran

Contoh perhitungan manual:

Evelyn Yoniana / 15522083

DBSCAN

$$x \text{ train} = \begin{bmatrix} 1 & 2 \\ 1 & 4 \\ 1 & 0 \\ 10 & 2 \\ 10 & 4 \\ 10 & 0 \end{bmatrix}$$

$\text{eps} = 3$ (maksimal jarak)

$\text{min sample} = 2$

$\text{metric} = \text{euclidean}$

jarak points ke $[1,2]$:	jarak points ke $[1,4]$:	jarak points ke $[10,2]$:	jarak points ke $[10,0]$:
$[1,2] : 0$	gawah dicari realnya	$[1,2] : 9$	gawah dicari realnya
$[1,4] : 2$	$[1,4]$ udah masuk cluster	$[1,4] : 9,22$	$[10,0]$ udah masuk cluster
$[1,0] : 2$		$[1,0] : 9,22$	
$[10,2] : 9$	jarak points ke $[1,0]$:	$[10,2] : 0$	jarak points ke $[10,0]$:
$[10,4] : 9,22$	gawah dicari realnya	$[10,4] : 2$	gawah dicari realnya
$[10,0] : 9,22$	$[1,0]$ udah masuk cluster	$[10,0] : 2$	$[10,0]$ udah masuk cluster
cluster 1 : $(1,2), (1,4), (1,0)$		cluster 2 : $(10,2), (10,4), (10,0)$	

Karena semua titik udah masuk cluster (kalo di codenya udah ganda yg labatnya -1), jadi perhitungannya stop di sini.