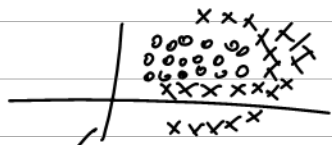


استفاده از فاصله بین نقاط برای clustering می تواند مشکلاتی را به همراه داشته باشد.

برای مثال اگر $decision\ boundary$ ما یک شکل غیر منظم بود نگاه به فاصله می تواند آن را



را هم جدا کند. به شکل دایره نگاه کنید.

در این مثال دو کلاس داریم که شکل غیر منظم دارند. فرض داریم که اگر حقیقتاً را به فاصله بگذاریم

به جواب اشتباه خواهیم رسید. همچنین نویز داده ها هم می تواند از دلایل عدم کاربرد فاصله باشد.

DBSCAN: این الگوریتم که بر اساس چگالی داده ها را جدا می کند دارای دو $hyperparameters$

هم است. یکی $min\ P_{db}$ و دیگری ϵ . برای هر نقطه یک دایره به شعاع ϵ رسم می کنیم. اگر

نقطه مورد نظر دیگر نقاط دایره $min\ P_{db}$ بود نگاه آن نقطه را متعلق به $core\ point$ می دانند.

و یک $neighbor$ می نامیم. همچنین برای هر نقطه انجام می دهیم. نقاطی که به $core\ point$ ها با

شعاع ϵ دسترسی دارند نیز به آن $neighbor$ اضافه می شوند. فقط نقاطی که در ϵ دایره آن

ها هیچ نقطه ای نیست به عنوان $Noise$ دسته بندی می شوند. باید این $min\ P_{db}$

چالش اصلی مسئله است. اما اگر بتوانیم مقادیر خوب را پیدا کنیم نگاه به چگالی می توان

شکل غیر منظم را هم $cluster$ کرد.

OPTICS: در الگوریتم $Optics$ ما دو فاصله را حساب می کنیم. یکی $reachability$ و دیگری $core$.

ابتدا نقاط را بر اساس فاصله اول که به فاصله $core\ point$ می نامند مرتب می کنیم. پس یک

$reachability\ plot$ می کشیم. در واقع نمودار اصلی در سمت راست این نقشه است. باز روی آن $cluster$

می کنند. از خوبی های آن این است که می خدای تاثیر گذار است و تفاوت چگالی مانند DBSCAN تاثیر خیلی

منفی نمی تواند بگذارد.