

## Dokumen Jawaban Unsupervised Learning

13520163/Frederik Imanuel Louis

1. Apa itu unsupervised learning?

Unsupervised Learning adalah algoritma yang dapat mencari pola dari suatu dataset tanpa intervensi manusia (unlabeled). Salah satu pola yang dapat dicari adalah pembentukan cluster dari data menggunakan algoritma clustering.

2. Jelaskan bagaimana cara kerja dari algoritma yang anda implementasikan!

DBScan memiliki parameter epsilon dan minimal points untuk menentukan apakah suatu titik membentuk cluster. Jika suatu titik memiliki titik sebanyak lebih dari minimal points yang berjarak maksimal epsilon dari titik tersebut, maka titik itu dan titik-titik tetangganya akan membentuk sebuah cluster. Kemudian, titik-titik tetangga tersebut akan diselidiki lagi apakah titik-titik tetangganya termasuk dalam cluster tersebut, sampai tidak ada titik lain yang masuk cluster tersebut. Titik-titik yang pada akhir pencarian tidak memiliki cluster diassign sebagai titik noise. Algoritma DBScan bersifat deterministik, dimana keluaran bergantung pada nilai eps dan minimal points.

KMeans secara acak memiliki k buah titik yang bertindak sebagai titik mean. Kemudian, seluruh titik akan diassign ke titik terdekat diantara k titik tersebut. Masing-masing titik mean tersebut akan membentuk sebuah cluster. Kemudian, titik mean akan diupdate dengan rata-rata semua titik yang berada pada cluster tersebut. Ini akan diulang sampai semua titik mean tidak berubah, atau iterasi mencapai maximum iteration. Karena pemilihan titik means awal dilakukan secara acak, maka KMeans bersifat nondeterministik.

KMedoids secara acak memilih k buah titik sebagai titik medoids awal. Kemudian, dihitung cost awal yaitu jumlah dari jarak/cost semua titik terhadap titik medoids terdekat dari titik tersebut. Clustering ini serupa dengan clustering awal KMeans. Kemudian, untuk masing-masing titik medoid, titik lain yang belum dipilih sebagai medoid akan ditukar dengan titik medoid tersebut, dan akan dihitung jumlah jarak/cost dengan titik baru tersebut sebagai medoid. Jika jumlah cost lebih kecil, maka titik baru tersebut akan bertindak sebagai medoid. Jika tidak, titik medoid awal akan ditukar kembali menjadi medoid. Karena pemilihan titik medoid awal dilakukan secara acak, maka KMedoids bersifat nondeterministik.

3. Bandingkan ketiga algoritma tersebut, kemudian tuliskan kelebihan dan kekurangannya!

KMedoids dan KMeans memaksa dataset menjadi K buah cluster, sedangkan DBScan membentuk cluster berdasarkan masukan jarak dan minimal point yang dilakukan, sehingga pembentukan cluster dapat lebih menggambarkan dataset dari algoritma lain. KMeans dan KMedoids membentuk cluster dengan cara yang serupa, tetapi KMeans membentuk cluster dengan lebih akurat karena iterasi dilakukan sampai titik means konvergen, sedangkan KMedoids membentuk cluster secara greedy. Secara kompleksitas waktu, KMedoids memiliki kompleksitas yang lebih rendah, sehingga lebih cocok digunakan untuk dataset besar, tetapi KMeans dapat memiliki execution time lebih cepat untuk dataset kecil karena titik means dapat konvergen lebih cepat dari KMedoids yang harus memilih semua titik medoid secara greedy.

4. Jelaskan penerapan dari algoritma unsupervised di dunia nyata!

Algoritma unsupervised, khususnya algoritma clustering, dapat digunakan untuk mencari pola dari dataset yang besar, misalnya memberi rekomendasi video youtube berdasarkan video lain yang sering dinonton suatu pengguna, dan google search yang menampilkan hasil pencarian menggunakan Search Result Clustering.