

TUGAS PENDAHULUAN PRAKTIKUM PERTEMUAN IX
PEMBELAJARAN MESIN



Oleh

Nama : *Muhammad Tijan A*

Nama : *Muhammad Reza Putra Anisi*

NIM : *105220023* NIM : *105220053*

PROGRAM STUDI ILMU KOMPUTER
FAKULTAS SAINS DAN KOMPUTER
UNIVERSITAS PERTAMINA
2022

Pertanyaan :

1. Jelaskan konsep dasar dari k-Nearest Neighbors (k-NN)!
2. Sebutkan dan jelaskan metode-metode pengukuran yang digunakan dalam k-NN!
3. Jelaskan secara rinci apa itu “k” dalam kNN! Apakah nilai dari k akan mempengaruhi model kNN? Jelaskan!

BONUS!!!

Sebutkan 3 varian algoritma yang digunakan untuk mempercepat proses perhitungan jarak pada kNN! Jelaskan!

Jawaban :

1. Konsep dasar dari k-Nearest Neighbors (k-NN) adalah sebuah metode dalam pembelajaran mesin yang digunakan untuk klasifikasi dan regresi. Metode ini didasarkan pada asumsi bahwa data dengan atribut yang mirip cenderung memiliki label atau nilai target yang sama. Berikut Langkah Langkah dasar

- Menentukan nilai K
- Menghitung jarak
- Memilih tetanga terdekat
- Menentukan label atau nilai target
- Menghasilkan hasil

Metode k-NN tidak melibatkan proses pelatihan yang kompleks seperti yang dilakukan oleh beberapa algoritma pembelajaran mesin lainnya. algoritma ini dapat membutuhkan waktu yang lama untuk memprediksi label atau nilai target pada dataset yang besar.

2. Metode pengukuran yang digunakan dalam algoritma k-NN (k-Nearest Neighbors) adalah metode untuk menghitung jarak atau kesamaan antara titik data yang sedang diklasifikasikan dengan tetangga terdekatnya

- Euclidean distance

Metode ini mengukur jarak antara dua titik dalam ruang Euclidean. Jarak Euclidean antara dua titik P dan Q dengan n atribut dihitung menggunakan rumus berikut:

$$D(P, Q) = \sqrt{(P_1 - Q_1)^2 + (P_2 - Q_2)^2 + \dots + (P_n - Q_n)^2}$$

- Manhattan distance

Metode ini juga dikenal sebagai jarak City Block atau jarak L1. Jarak Manhattan antara dua titik P dan Q dengan n atribut dihitung menggunakan rumus berikut:

$$D(P, Q) = |P_1 - Q_1| + |P_2 - Q_2| + \dots + |P_n - Q_n|$$

- Minkowski distance

Metode ini adalah generalisasi dari jarak Euclidean dan Manhattan. Jarak Minkowski antara dua titik P dan Q dengan n

- Hamming distance

Metode ini digunakan khusus untuk data dengan atribut biner (0 dan 1).

Jarak Hamming antara dua titik P dan Q dengan n atribut dihitung sebagai jumlah posisi di mana nilai atribut keduanya berbeda

Metode pengukuran jarak ini memungkinkan algoritma k-NN untuk membandingkan dan mengklasifikasikan data baru berdasarkan jarak atau kesamaan mereka dengan tetangga terdekat dalam ruang atribut

3. Dalam algoritma k-NN (k-Nearest Neighbors), "k" merujuk pada jumlah tetangga terdekat yang akan digunakan untuk melakukan klasifikasi atau prediksi pada data yang sedang dievaluasi. Dengan kata lain, "k" menentukan berapa banyak tetangga terdekat yang akan mempengaruhi hasil klasifikasi atau prediksi.

Berikut adalah beberapa hal yang perlu diperhatikan terkait pemilihan nilai k:

- Underfitting dan Overfitting: Jika nilai k terlalu kecil, seperti $k = 1$, model akan cenderung overfitting. Ini berarti model akan sangat responsif terhadap variabilitas kecil dalam data pelatihan, yang dapat menyebabkan klasifikasi yang tidak stabil atau noise.
- Bias dan Variansi: Nilai k juga mempengaruhi trade-off antara bias dan variansi dalam model k-NN. Nilai k yang kecil cenderung memiliki bias yang lebih tinggi, karena hanya sedikit tetangga yang dipertimbangkan, dan dapat menghasilkan prediksi yang lebih sederhana
- Keseimbangan dan Ukuran Dataset: Pemilihan nilai k juga harus mempertimbangkan keseimbangan antara kompleksitas model dan ukuran dataset.

BONUS :

- Algoritma KNN adaptif adalah varian yang berfokus pada pemilihan nilai k optimal untuk data. Ini bekerja dengan menerapkan algoritma terpisah untuk menentukan nilai k optimal untuk setiap titik data dari dataset pelatihan.
- Algoritma fuzzy KNN berputar di sekitar prinsip penugasan keanggotaan. Mirip dengan algoritme KNN klasik, varian melanjutkan untuk menemukan k tetangga terdekat dari kumpulan data pengujian dari kumpulan data pelatihan.
- Algoritma Hassanat KNN merupakan varian yang memiliki titik fokus pada rumus pengukuran jarak. Varian ini mengikuti desain sederhana dari algoritma KNN, tetapi menawarkan cara lanjutan untuk menemukan jarak antara dua titik data.

Referensi :

<https://www.analyticsvidhya.com/blog/2021/08/how-knn-uses-distance-measures/>

<https://www-nature-com.translate.goog/articles/s41598-022-10358->

x?error=cookies_not_supported&code=fb16a1ac-ada6-4e0a-8ba2-

e5510f136030&x_tr_sl=en&x_tr_tl=id&x_tr_hl=id&x_tr_pto=tc

<https://www.ibm.com/topics/knn>

<https://www.analyticsvidhya.com/blog/2021/08/how-knn-uses-distance-measures/>