

A few thoughts on picking a value for “K”

- There is no physical or biological way to determine the best value for “K”, so you may have to try out a few values before settling on one. Do this by pretending part of the training data is “unknown”.

⏮ ⏪ 🔊 4:41 / 5:30 • Thoughts on how to pick 'K' >

Pencarian langkah kaca itu halo dan selamat datang di pencarian langkah tumpukan pencarian dipersembahkan oleh orang-orang yang ramah di departemen genetika di Universitas North Carolina di Chapel Hill hari ini kita akan berbicara tentang K terdekat algoritma tetangga yang merupakan cara super sederhana untuk mengklasifikasikan data secara singkat jika Anda sudah memiliki banyak data yang mendefinisikan tipe sel ini, kita dapat menggunakannya untuk memutuskan tipe sel yang mana orang ini, mari kita lihat tindakannya langkah 1 mulai dengan a kumpulan data dengan kategori yang diketahui dalam hal ini kami memiliki tipe sel yang berbeda dari tumor usus kami kemudian mengelompokkan data tersebut dalam hal ini kami menggunakan PCA langkah 2 menambahkan sel baru dengan kategori yang tidak diketahui ke bekuan darah kami tidak mengetahui kategori sel ini karena itu diambil dari tumor lain yang sel-selnya tidak diurutkan dengan benar, jadi apa yang ingin kita lakukan adalah mengklasifikasikan sel baru ini.

a) K-Tetangga Terdekat (K-Nearest Neighbors atau K-NN):

K-Nearest Neighbors (K-NN) adalah algoritma klasifikasi dan regresi sederhana yang digunakan dalam machine learning. Ini beroperasi berdasarkan asumsi bahwa jika sekelompok data mirip dalam fitur-fiturnya, maka mereka lebih mungkin memiliki label yang sama (klasifikasi) atau nilai target yang mirip (regresi).

- Pada dasarnya, K-NN bekerja dengan menghitung jarak antara titik data yang akan diprediksi dan titik data lainnya dalam dataset.
- "K" dalam K-NN mengacu pada jumlah tetangga terdekat yang akan digunakan untuk membuat prediksi.
- Misalnya, jika Anda memilih $K=3$, algoritma akan mencari 3 tetangga terdekat yang memiliki fitur-fitur serupa dengan titik data yang akan diprediksi.
- K-NN akan memutuskan label atau nilai target untuk titik data tersebut berdasarkan mayoritas label tetangga-tetangga tersebut. Jika, misalnya, 2 dari 3 tetangga terdekat memiliki label "A" dan 1 memiliki label "B", maka K-NN akan memprediksi label "A" untuk titik data tersebut.

Dalam penggunaan praktis, penting untuk memilih nilai K yang tepat. Nilai K yang terlalu kecil dapat membuat model rentan terhadap noise dalam data, sementara nilai K yang terlalu besar dapat membuat model terlalu kasar.

K -NN adalah salah satu algoritma yang mudah dimengerti dan sering digunakan, terutama untuk masalah klasifikasi sederhana. Ini cocok untuk dataset dengan jumlah sampel yang tidak terlalu besar, tetapi mungkin tidak efisien untuk data dengan banyak fitur atau titik data dalam jumlah besar.