

Berikan ringkasan dari artikel berikut ini

<https://www.pyimagesearch.com/2016/08/08/k-nn-classifier-for-imageclassification/>

JAWAB

Kumpulan data Kaggle Dogs vs. Cats

Kumpulan data Anjing vs. Kucing sebenarnya adalah bagian dari tantangan Kaggle beberapa tahun yang lalu. Tantangannya sendiri sederhana: diberi gambar, memprediksi apakah itu berisi anjing atau kucing. Cukup sederhana — tetapi jika Anda mengetahui sesuatu tentang klasifikasi gambar, Anda akan memahaminya karena:

- Variasi sudut pandang
- Variasi skala
- Deformasi
- Halangan
- Kekacauan latar belakang
- Variasi intra-kelas

Bahwa masalahnya secara signifikan lebih sulit daripada yang mungkin muncul di permukaan.

Dengan hanya menebak secara acak, Anda harus dapat mencapai akurasi 50% (karena hanya ada dua label kelas). Algoritme pembelajaran mesin perlu mendapatkan akurasi > 50% untuk menunjukkan bahwa ia sebenarnya telah "mempelajari" sesuatu (atau menemukan pola yang mendasari dalam data).

Bagaimana cara kerja pengklasifikasi k-NN?

K-Nearest Neighbor classifier sejauh ini merupakan algoritma pembelajaran mesin/klasifikasi gambar yang paling sederhana. Faktanya, ini sangat sederhana sehingga tidak benar-benar "mempelajari" apa pun.

Di dalam, algoritme ini hanya mengandalkan jarak antara vektor fitur, seperti membangun mesin pencari gambar — hanya saja kali ini, kami memiliki label yang terkait dengan setiap gambar sehingga kami dapat memprediksi dan mengembalikan kategori sebenarnya untuk gambar tersebut.

Sederhananya, algoritma k-NN mengklasifikasikan titik data yang tidak diketahui dengan mencari kelas yang paling umum di antara contoh k-terdekat. Setiap titik data dalam k contoh terdekat memberikan suara dan kategori dengan suara terbanyak menang!

Atau, dalam bahasa Inggris sederhana: "Katakan siapa tetangga Anda, dan saya akan memberi tahu Anda siapa Anda"

Untuk memvisualisasikannya, lihat contoh mainan berikut di mana saya telah memplot "kelembutan" hewan di sepanjang sumbu x dan ringannya bulu mereka pada sumbu y:

Jelaskan pengertian Euclidean distance pada artikel tersebut

JAWAB

Dalam matematika, jarak Euclidean antara dua titik dalam ruang Euclidean adalah panjang segmen garis antara dua titik. Ini dapat dihitung dari koordinat Cartesian dari titik-titik menggunakan teorema Pythagoras, oleh karena itu kadang-kadang disebut jarak Pythagoras. Nama-nama ini berasal dari matematikawan Yunani kuno Euclid dan Pythagoras, meskipun Euclid tidak mewakili jarak sebagai angka, dan hubungan dari teorema Pythagoras dengan perhitungan jarak baru dibuat pada abad ke-18.

Jarak antara dua benda yang bukan titik biasanya didefinisikan sebagai jarak terkecil antara pasangan titik dari dua benda. Rumus dikenal untuk menghitung jarak antara berbagai jenis objek, seperti jarak dari titik ke garis. Dalam matematika tingkat lanjut, konsep jarak telah digeneralisasikan ke ruang metrik abstrak, dan jarak selain Euclidean telah dipelajari. Dalam beberapa aplikasi dalam statistik dan optimasi, kuadrat jarak Euclidean digunakan sebagai pengganti jarak itu sendiri.