

UNDERSTANDING 3 LINK STATQUEST

Nama : Al Ghifary Akmal Nasheeri

NIM : 1103201242

Kelas : TK-44-06

StatQuest: Principal Component Analysis (PCA), Step-by-Step by Josh Starmer

1. Pengenalan PCA

PCA adalah teknik analisis data yang digunakan untuk mengurangi dimensi variabel dalam data. Tujuannya adalah untuk mengidentifikasi pola-pola penting dalam data dan mengurangi kompleksitas data.

2. Standarisasi

Sebelum menerapkan PCA, penting untuk melakukan standarisasi data, yaitu mengubah skala variabel sehingga memiliki rata-rata 0 dan simpangan baku 1. Hal ini memastikan semua variabel berada pada skala yang sama.

3. Matriks Kovarians

PCA bekerja dengan membuat matriks kovarians, yang menggambarkan hubungan antar variabel dalam data. Matriks ini menunjukkan bagaimana variabel-variabel berubah bersamaan.

4. Vektor Eigen dan Nilai Eigen

PCA didasarkan pada vektor eigen dan nilai eigen matriks kovarians. Vektor eigen menggambarkan arah di mana data bervariasi paling banyak, sedangkan nilai eigen mengindikasikan seberapa banyak variasi yang dijelaskan oleh vektor eigen tersebut.

5. Pengurutan Vektor Eigen

Vektor eigen diurutkan berdasarkan nilai eigen yang sesuai, dengan vektor eigen terbesar menggambarkan komponen utama pertama, yang terbesar kedua, dan seterusnya.

6. Membuat Komponen Utama

Komponen utama adalah kombinasi linear dari variabel asli yang tidak berkorelasi satu sama lain dan menangkap variasi data. Komponen utama pertama menangkap variasi paling banyak, yang kedua menangkap yang kedua paling banyak, dan seterusnya.

7. Reduksi Dimensi

Fitur untuk memilih dalam menyimpan sejumlah komponen utama untuk mengurangi dimensi data. Dengan cara ini, data dapat direpresentasikan dengan lebih sedikit variabel sambil mempertahankan sebagian besar informasi.

8. Interpretasi

Komponen utama mungkin tidak memiliki interpretasi fisik langsung, tetapi mereka merepresentasikan pola atau kombinasi variabel asli yang memaksimalkan variasi.

9. Aplikasi

PCA digunakan dalam berbagai bidang seperti analisis genetik, pemrosesan gambar, keuangan, dan pembelajaran mesin.

StatQuest: K-nearest neighbors, Clearly Explained by Josh Starmer

1. Pengenalan K-Nearest Neighbors (K-NN)

Algoritma K-NN adalah salah satu algoritma pembelajaran mesin yang digunakan dalam masalah klasifikasi dan regresi. Ini adalah algoritma yang berbasis pada "belajar dari tetangga terdekat."

2. Konsep Dasar

Algoritma K-NN bekerja dengan cara membandingkan data yang akan diprediksi dengan data pelatihan yang sudah ada. Ini mengukur seberapa mirip data yang akan diprediksi dengan data pelatihan yang ada.

3. Parameter K

Parameter K dalam K-NN adalah jumlah tetangga terdekat yang akan digunakan untuk membuat prediksi. Memilih nilai K yang tepat sangat penting, dan berbagai nilai K dapat menghasilkan hasil yang berbeda.

4. Langkah-Langkah K-NN

- Hitung jarak antara data yang akan diprediksi dan setiap titik data pelatihan.
- Pilih K titik data pelatihan terdekat.
- Lakukan mayoritas suara (dalam klasifikasi) atau perhitungan rata-rata (dalam regresi) dari K tetangga terdekat ini untuk membuat prediksi.

5. Klasifikasi vs. Regresi

K-NN dapat digunakan baik untuk masalah klasifikasi (prediksi label kategori) maupun regresi (prediksi nilai kontinu).

6. Parameter Jarak

Metrik jarak yang digunakan untuk mengukur seberapa mirip data bisa beragam, misalnya Euclidean, Manhattan, atau metrik jarak lainnya.

7. Kelebihan dan Keterbatasan

- Kelebihan: Mudah dimengerti, bekerja baik dengan dataset kecil, tidak memerlukan pembelajaran berat.

- Keterbatasan: Sensitif terhadap nilai K yang dipilih, tidak efisien untuk dataset besar, dan dapat terpengaruh oleh data pencilan.

8. Aplikasi

K-NN digunakan dalam berbagai bidang, termasuk pengenalan pola, klasifikasi teks, rekomendasi produk, dan pengenalan wajah.