

Nama : Daffa Asyqar Ahmad Khalisheka

NIM : 1103200034

Kelas : TK-44-G4

- Principal Component Analysis

Analisis Komponen Utama, juga dikenal sebagai PCA, adalah metode statistik yang digunakan untuk mengidentifikasi pola dasar dalam data dengan mengurangi dimensinya. Banyak bidang, termasuk statistik, ilmu komputer, ilmu sosial, dan ekonomi, menggunakan PCA untuk analisis, pengurangan dimensi, dan visualisasi data. Berikut adalah tahapan utama PCA:

1. Standardisasi Data: Untuk memastikan bahwa setiap variabel memberikan kontribusi yang sama untuk analisis, PCA melakukan langkah pertama dengan mengukur data dalam skala yang sama, yaitu dengan menghilangkan rata-rata dari setiap fitur dan membaginya dengan deviasi standar.
2. Perhitungan Matriks Kovariansi: Selanjutnya, kita mengambil data yang telah distandarisasi dan menghitung matriks kovariansi. Matriks kovariansi menghitung bagaimana setiap pasangan variabel dalam data berhubungan satu sama lain.
3. Menghitung Vektor Eigen dan Nilai Eigen: Penghitungan vektor eigen dan nilai eigen dari matriks kovariansi dapat dilakukan dengan PCA. Nilai eigen menunjukkan seberapa penting setiap vektor eigen untuk menjelaskan variasi data.
4. Pemilihan Komponen Utama: Komponen utama adalah arah utama variasi data dan adalah vektor eigen yang sesuai dengan nilai eigen terbesar.
5. Proyeksi Data: Selanjutnya, data asli diproyeksikan ke ruang yang dibentuk oleh komponen utama yang dipilih. Dataset dimensi yang lebih rendah diciptakan sebagai hasil dari proyeksi ini.
6. Interpretasi Hasil: Hasil PCA dapat ditafsirkan sesuai dengan masalah yang sedang dianalisis. Dalam domain masalah, komponen utama yang dipilih biasanya memiliki makna.

Salah satu manfaat PCA adalah kemampuan untuk memvisualisasikan pola dalam data, mengurangi dimensinya, dan mengidentifikasi atribut yang paling penting. Namun, perlu diingat bahwa PCA adalah teknik linier, sehingga tidak dapat menangani pola data yang lebih kompleks. Selain itu, dalam praktiknya, asumsi PCA bahwa komponen utama bersifat ortogonal (tidak berkorelasi satu sama lain) tidak selalu benar.

Sangat penting untuk memahami bahwa PCA adalah teknik yang umum digunakan untuk analisis data numerik. PCA dapat digunakan untuk data non-numerik. Analisis Komponen Utama Berbasis

Korelasi (Principal Component Analysis based on Correlation) dan Analisis Komponen Utama Berbasis Peringkat (Principal Component Analysis based on Rankings) adalah dua contoh dari berbagai jenis data yang dapat digunakan PCA.

- K-nearest neighbors

K-Nearest Neighbors (K-NN) adalah salah satu algoritma pengajaran mesin yang digunakan untuk tugas regresi dan klasifikasi. Prinsip dasar algoritma ini adalah bahwa objek yang mirip cenderung berada di dekat satu sama lain dalam ruang fitur. Konsep utama K-NN adalah mencari K titik data terdekat dari dataset pelatihan, di mana K adalah sejumlah tetap yang ditentukan sebelumnya. Ini digunakan untuk mengklasifikasikan atau memprediksi titik data. Algoritma kemudian menentukan kelas atau nilai target yang akan diberikan kepada titik data yang diprediksi dengan melihat sebagian besar kelas dari tetangga terdekat. K-NN sangat sederhana secara konseptual dan tidak membutuhkan pelatihan model yang kompleks. Namun, karena perlu menghitung jarak antara titik data yang akan diprediksi dengan seluruh titik data dalam dataset pelatihan, itu memiliki kelemahan dalam hal komputasi. Selain itu, hasil prediksi dapat dipengaruhi oleh pemilihan parameter K. Oleh karena itu, K-NN terus menjadi alat yang bermanfaat untuk berbagai tujuan, terutama dalam kasus di mana dataset kecil dan mudah diinterpretasikan.

- Decision Tree

Salah satu teknik pengolahan data dan pengajaran mesin yang digunakan untuk pengambilan keputusan. Ini adalah model prediktif yang menunjukkan hierarki keputusan dan konsekuensi yang mungkin terjadi. Pohon keputusan ini terdiri dari "simpul" atau "daun" yang menunjukkan tes pada atribut data, "cabang" yang menunjukkan hasil tes, dan "daun" yang menunjukkan nilai prediksi atau keputusan akhir. Proses pembuatan pohon keputusan dimulai dengan memilih fitur terbaik untuk berfungsi sebagai akar pohon, yang merupakan fitur yang paling informatif dalam memisahkan data. Nilai-nilai fitur tersebut kemudian digunakan untuk membagi pohon menjadi cabang yang lebih kecil. Proses ini berlanjut hingga semua data dimasukkan ke dalam daun-daun yang sesuai dengan keputusan atau prediksi akhir. Klasifikasi, regresi, atau masalah pengambilan keputusan lainnya dapat diatasi dengan pohon keputusan, dan mereka membantu kita memahami hubungan antara atribut dan hasil dalam data. Selain itu, decision tree mudah dipahami dan digunakan dalam berbagai bidang, seperti teknologi, kesehatan, dan bisnis, dengan tujuan meningkatkan kemampuan individu untuk membuat keputusan yang lebih cerdas.