

Nama : Mohamad Rayi Dwi Putra

NIM : 1103213070

Task 5 Membuat Catatan PCA

Analisis Komponen Utama (PCA) adalah teknik statistik yang digunakan untuk mereduksi dimensi data tanpa kehilangan informasi penting. Teknik ini mengidentifikasi komponen utama dalam data, yang merupakan kombinasi linear dari variabel asli. PCA (Principal Component Analysis) dapat bekerja pada data 2 dimensi dengan cara:

1. Standarisasi data: Pastikan semua variabel memiliki skala yang sama dengan menstandarisasinya (mengurangi mean dan membaginya dengan standar deviasi).
2. Hitung matriks kovarians: Matriks kovarians menunjukkan hubungan antara setiap pasangan variabel. Dalam data 2 dimensi, matriks kovarians adalah matriks 2×2 .
3. Hitung nilai eigen dan eigenvektor dari matriks kovarians:
 - Nilai eigen: menunjukkan variabilitas yang dijelaskan oleh masing-masing komponen utama. Dalam data 2 dimensi, terdapat dua nilai eigen.
 - Eigenvektor: menunjukkan arah komponen utama. Dalam data 2 dimensi, terdapat dua eigenvektor, yang direpresentasikan sebagai vektor 2 dimensi.
4. Pilih komponen utama: Pilih komponen utama yang menjelaskan proporsi variabilitas yang cukup. Dalam data 2 dimensi, biasanya hanya satu komponen utama yang dipilih, yaitu yang memiliki nilai eigen lebih besar.
5. Transformasi data: Transformasikan data ke ruang komponen utama. Dalam data 2 dimensi, transformasi ini menghasilkan satu variabel baru yang merupakan kombinasi linear dari dua variabel asli.

Motivasi Konseptual untuk Analisis Komponen Utama (PCA)

1. Mengatasi Kompleksitas Data:

- Dunia nyata dipenuhi dengan data yang kompleks, seringkali dengan banyak variabel yang saling terkait.
- Mengelola dan memahami data ini bisa menjadi rumit.
- PCA menawarkan solusi dengan mereduksi dimensi data tanpa kehilangan informasi penting.

2. Menemukan Struktur Data:

- Data sering kali memiliki struktur yang tersembunyi di balik banyak variabel.
- PCA membantu mengidentifikasi struktur ini dengan menemukan komponen utama yang menjelaskan sebagian besar variabilitas data.

3. Meningkatkan Interpretabilitas:

- Data dengan dimensi tinggi sulit untuk diinterpretasikan.
- PCA menghasilkan variabel baru (komponen utama) yang lebih mudah diinterpretasikan.

4. Visualisasi Data:

- Data dengan dimensi tinggi sulit untuk divisualisasikan.
- PCA membantu memvisualisasi data dengan mereduksinya ke dimensi yang lebih rendah.

5. Efisiensi Komputasi:

- Algoritma dan model statistik sering kali lebih efisien dengan data berdimensi rendah.
- PCA membantu meningkatkan efisiensi komputasi dengan mereduksi dimensi data.

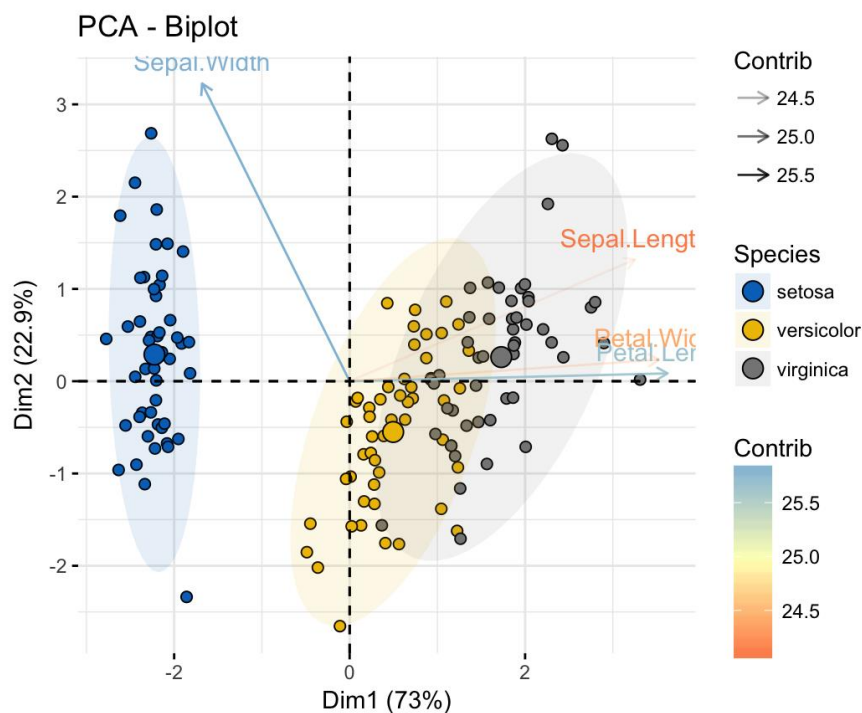
6. Mengurangi Kebisingan:

- Data often contains noise that can obscure important patterns.
- PCA helps to reduce noise by focusing on the principal components that capture the most significant information.

7. Menemukan Hubungan Antar Variabel:

- PCA membantu menemukan hubungan antar variabel dengan menganalisis eigenvektor.
- Eigenvektor menunjukkan arah dan kekuatan hubungan antar variabel.

Menggambar Grafik PCA



Grafik PCA yang paling umum digunakan adalah scatter plot dari komponen utama (PC) yang dihasilkan.

Berikut langkah-langkah untuk menggambar grafik PCA:

1. Hitung Komponen Utama: Setelah Anda menerapkan PCA pada data Anda, Anda akan memiliki skor komponen utama untuk setiap data point. Skor ini mewakili koordinat data point dalam ruang komponen utama.
2. Pilih Komponen Utama: Biasanya, Anda akan memplot dua komponen utama pertama (PC1 dan PC2) karena mereka menjelaskan sebagian besar variabilitas data.
3. Buat Scatter Plot: Plotkan skor PC1 pada sumbu x dan skor PC2 pada sumbu y. Setiap data point akan direpresentasikan sebagai titik pada scatter plot.
4. Interpretasi: Analisis posisi titik-titik pada scatter plot untuk melihat bagaimana data dikelompokkan atau tersebar. Pola dan struktur dalam data akan terungkap dalam scatter plot PCA.