

NAMA : APRIL HAMONANGAN MARBUN

NIM : 1103202039

PCA

Principal Component Analysis (PCA) adalah metode reduksi dimensi yang ampuh yang diterapkan secara luas dalam Machine Learning dan analisis data. Dengan memaksimalkan varians data di sepanjang komponen utama, analisis komponen utama (PCA) mencoba mengubah data berdimensi tinggi menjadi sistem koordinat baru. Hal ini memungkinkan pengurangan noise dan penyajian data dengan mengurangi ukuran fitur tanpa mengorbankan informasi penting. PCA digunakan secara luas dalam aplikasi seperti pengurangan gambar dan pengenalan pola dalam kumpulan data yang besar dan rumit. PCA juga sangat berguna untuk mengungkap struktur tersembunyi dalam data dan meningkatkan model pembelajaran mesin. Saya telah menggunakan PCA secara efektif untuk mengoptimalkan sistem rekomendasi, meningkatkan kepuasan pengguna, dan menghasilkan pendapatan. Selain itu, PCA menyederhanakan penanganan dan pengarsipan data diagnostik dalam analisis citra medis.

-Analisis Komponen Utama (PCA) - Penjelasan Sederhana:

PCA seperti mengambil kumpulan data yang besar dan kompleks dan menyederhanakannya sehingga kita dapat memahaminya dengan lebih baik. Bayangkan Anda memiliki data tentang tinggi dan berat badan seseorang, dan Anda ingin menemukan pola atau mengurangi informasi tersebut menjadi beberapa ide utama.

-Memusatkan Data:

Pertama, kita cari rata-rata tinggi dan berat badan. Kemudian, kita sesuaikan semua titik data sehingga rata-rata berada di tengah. Hal ini membantu kita fokus pada bagaimana setiap orang berbeda dari rata-rata.

-Menemukan Hubungan:

Kami memeriksa apakah tinggi dan berat badan berhubungan. Jika ya, kami ingin tahu seberapa besar hubungannya. Jika tidak terlalu berhubungan, kita berada di jalur yang benar karena keduanya memberikan informasi yang berbeda.

-Memilih Info Terbaik:

PCA mengidentifikasi cara "terbaik" untuk menggabungkan informasi tinggi dan berat badan untuk menangkap perbedaan yang paling penting dalam data. Kombinasi baru ini adalah "komponen utama" kami.

-Menyederhanakan Data:

Kita dapat memutuskan untuk menyimpan komponen utama yang paling penting saja (yang menjelaskan variasi paling banyak), atau kita dapat menyimpan lebih banyak komponen untuk mendapatkan gambaran yang lebih kaya.

-Menggunakan Data yang Disederhanakan:

Sekarang kita memiliki versi yang lebih sederhana dari data kita, yang masih memberi tahu kita hal-hal penting yang perlu kita ketahui. Ini seperti meringkas sebuah cerita yang panjang menjadi beberapa poin penting.

LDA

Model probabilistik mendasar yang digunakan dalam pemodelan topik dan pemrosesan bahasa alami disebut Latent Dirichlet Allocation (LDA). Model ini menyatakan bahwa dokumen adalah kumpulan subjek, yang masing-masing terdiri dari kata-kata. Tujuan utama dari LDA adalah untuk mengidentifikasi tema-tema laten di dalam kumpulan data dan distribusi kata-katanya. LDA memodelkan probabilitas ini menggunakan distribusi Dirichlet dan menggunakan teknik inferensi untuk menyimpulkan distribusi topik dalam dokumen. Analisis sentimen, peringkasan subjek, rekomendasi konten, dan kategorisasi dokumen adalah beberapa di antara sekian banyak aplikasi LDA. Namun, mungkin sulit untuk memutuskan berapa banyak subjek yang ideal dan untuk memastikan bahwa informasinya dapat dipahami. Untuk mendapatkan temuan yang relevan, persiapan data yang efektif juga sangat penting. Terlepas dari tantangan-tantangan ini, LDA terus menjadi metode yang efektif untuk mengungkap pola-pola tersembunyi dalam data teks, memberikan informasi yang mendalam di berbagai bidang.

Latent Dirichlet Allocation (LDA) - Penjelasan Sederhana:

LDA seperti alat ajaib untuk memahami setumpuk dokumen atau artikel. Bayangkan Anda memiliki sekotak cerita, dan Anda ingin tahu apa isi cerita-cerita ini tanpa membaca semuanya.

-Mengurutkan Kata:

LDA dimulai dengan menyortir semua kata dalam cerita ke dalam kategori yang berbeda. Kategori-kategori ini seperti topik. Setiap topik mewakili sekelompok kata yang berhubungan.

-Mencampurkan Topik:

Sekarang, LDA menganggap setiap cerita sebagai campuran dari topik-topik ini. Ini seperti mengatakan, "Cerita ini sebagian besar tentang topik A, sedikit tentang topik B, dan sedikit tentang topik C"

-Matematika Ajaib:

LDA menggunakan matematika yang cerdas untuk mengetahui topik mana yang ada di setiap cerita dan seberapa kuat topik tersebut. Ini seperti menemukan resep untuk setiap cerita dengan jumlah yang tepat untuk setiap topik.

-Menemukan Topik:

Setelah melakukan hal ini untuk semua cerita, LDA memberi tahu Anda apa topik utamanya. Anda dapat menganggap topik-topik ini sebagai ide-ide besar yang mengalir di semua cerita.

-Membuat Makna: (Making Sense)

LDA membantu Anda memahami apa yang ada di dalam tumpukan dokumen Anda. Anda dapat menggunakannya untuk melihat apakah cerita Anda tentang hewan, makanan, perjalanan, atau topik lainnya, bahkan jika Anda memiliki banyak sekali tumpukan dokumen. Ini seperti memiliki pustakawan super cepat yang mengkategorikan semua buku Anda tanpa membaca setiap halaman.

SVD

Singular Value Decomposition (SVD) adalah metode faktorisasi matriks dasar yang digunakan dalam sistem rekomendasi, analisis data, dan reduksi dimensi. Sebuah matriks diuraikan menjadi tiga bagian: U , Σ , dan V . Vektor singular kiri diwakili oleh U , vektor singular kanan dimasukkan ke dalam V , dan nilai singular ditempatkan di matriks diagonal Σ . SVD memainkan peran penting dalam pengurangan dimensi dengan menyederhanakan data yang rumit dengan hanya menyimpan nilai singular tertinggi dan vektor yang berhubungan dengannya. Aplikasi untuk teknologi ini termasuk penyaringan kolaboratif dalam sistem rekomendasi, analisis semantik laten untuk penggalian teks, dan kompresi gambar. Namun, karena masalah biaya komputasi dan penskalaan, SVD tidak lagi relevan dalam banyak pekerjaan berbasis data dan sekarang digunakan dengan pendekatan seperti Truncated SVD.

Singular Value Decomposition (SVD) - Penjelasan Sederhana:

SVD seperti alat khusus untuk menemukan cerita tersembunyi dalam tabel angka yang besar, seperti spreadsheet atau database. Bayangkan Anda memiliki sebuah tabel dengan banyak data, dan Anda ingin melihat bagian yang paling penting dari tabel tersebut.

-Memecahnya

SVD mengambil tabel besar ini dan memecahnya menjadi tiga bagian yang lebih kecil. Anggap saja sebagai potongan puzzle: satu bagian memiliki arah khusus (seperti Utara, Selatan, Timur, dan Barat), satu bagian memiliki angka-angka penting, dan bagian terakhir memiliki lebih banyak arah khusus.

-Angka-angka Ajaib

Bidak dengan angka penting memberi tahu Anda seberapa besar setiap arah dan seberapa penting arah tersebut. Ini seperti mengatakan, "Utara dua kali lebih penting daripada Timur, dan Selatan tiga kali lebih penting daripada Barat."

-Melihat Gambaran Besar

Dengan menggunakan potongan-potongan teka-teki ini, SVD memungkinkan Anda memahami pola-pola utama dalam data Anda. Hal ini membantu Anda mengetahui apa yang penting dan apa yang tidak penting. Ini seperti memiliki peta yang menunjukkan kepada Anda tempat-tempat paling menarik untuk dikunjungi dalam perjalanan.

-Penyederhanaan Data

Anda dapat menggunakan SVD untuk menyederhanakan data Anda. Misalnya, jika Anda memiliki banyak data dengan banyak detail, SVD dapat membantu Anda fokus pada aspek yang paling penting dan mengabaikan hal-hal yang kurang penting.

-Aplikasi

Orang-orang menggunakan SVD dalam berbagai hal, seperti kompresi gambar, membuat rekomendasi yang lebih baik di situs web, dan bahkan memahami hubungan antara kata-kata dalam bahasa alami. Ini adalah alat yang berguna untuk mengungkap