

Nama : Dena Cahya Setia Putri

Kelas : JumaTec

Statistical Test & Dimensionality Reduction

Dimensionality reduction atau reduksi dimensi adalah teknik untuk mengurangi dimensi dataset yang dalam hal ini adalah fitur data. Biasanya dataset yang akan diproses memiliki puluhan bahkan mungkin ratusan fitur atau kolom, maka dengan reduksi dimensi kita dapat mengurangi jumlah fitur atau kolom tanpa menghilangkan informasi dari dataset tersebut.

1. PCA (Principal component analysis) teknik yang paling populer. Teknik ini menggunakan operasi matriks sederhana dari aljabar linier dan statistik untuk menghiyung proyeksi dari data asli ke dalam dimensi dengan jumlah yang sama atau lebih sedikit.
2. LDA (linear discriminant analysis) merupakan salah satu algoritma ML untuk klasifikasi dapat digunakan untuk Dimensionality reduction. Cara menghitungnya dg menghitung ringkasan statistik untuk feature2 input menurut labelX seperti mean dan standard deviation. Menurut para peneliti DS Tharwat Alaa: LDA adalah teknik yang sangat umum untuk masalah pengurangan dimensi sebagai langkah pra pemrosesan untuk pembelajaran mesin dan aplikasi klasifikasi pola. LDA adalah metode mentransformasi data menjadi feature baru yang memperbesar jarak antar kelas. Secara umum prosesnya mirip dengan PCA. Persamaan PCA & LDA: sama sama dihitung dengan nilai eigen. Perbedaanya: PCA: lebih menghitung atau menjadi variansnya

Self Learning

Pada tugas kali ini kami akan menguji/membuat suatu program untuk *statistical class* dan dimensi reduksi melalui dataset yang telah disediakan dengan algoritma KNN yang akan menguji perbedaan hasil dari PCA dan LDA. Proses coding dapat dilihat melalui tautan berikut:

https://colab.research.google.com/drive/1dRfuHtKaGNFWf09Z6_zPPX1E0qVPD49?usp=s_haring

Dari proses tersebut maka dapat ditarik kesimpulan bahwa jumlah data yang banyak dan data yang lebih sedikit menghasilkan akurasi yang sama dan namun perbedaanya terletak pada waktu yang dibutuhkan untuk memprosesnya, jelas data yang lebih sedikit akan lebih cepat, dan sebaliknya.