

## ***Principal Component Analysis***

1. Jelaskan cara kerja dari algoritma tersebut! (boleh dalam bentuk *pseudocode* ataupun narasi)

⇒ Pertama, model menyimpan jumlah komponen paling berpengaruh yang ingin diambil. Kemudian model juga menyimpan data masukan dan menghitung rata-rata nilai dari untuk setiap fitur data masukan tersebut, nilai rata-rata tersebut kemudian akan digunakan untuk menempatkan data agar berada di tengah-tengah bidang kartesian dengan mengurangi nilai pada data dengan rata-rata data tersebut. Setelah itu dilakukan perhitungan kovarian matriks dari data masukan yang telah dilakukan *transpose*. Setelah mendapatkan kovarians matriks, selanjutnya mendapatkan nilai *eigenvector* dan *eigenvalue* dengan menggunakan fungsi bawaan *numpy* :

```
np.linalg.eig(cov)
cov : kovarian matriks
```

Setelah itu nilai *eigenvalue* diurut mulai dari yang terbesar, tetapi yang disimpan adalah berupa index pada data, jadi index data yang memiliki nilai *eigenvalue* terbesar akan berada di paling depan. Disini dilakukan pengurutan menurun pada nilai *eigenvalue* karena nilai *eigenvalue* disini merupakan nilai variansi data, jadi semakin besar nilai variansi yang didapatkan artinya kita bisa mendapatkan banyak informasi dari data tersebut yang memiliki variansi yang besar. Pada akhirnya diambil sebanyak *n\_component* pertama yang memiliki nilai *eigenvalue* terbesar.

2. Bandingkan hasil evaluasi model from scratch dan *library*, bagaimana hasil perbandingannya? Jika ada perbedaan, jelaskan alasannya!

Waktu Eksekusi	
<i>Scratch</i>	<i>Library</i>
0.002 detik	0.013 detik
<i>Explained Variance Ratio</i>	
<i>Scratch</i>	<i>Library</i>
[PC1] 0.36881	[PC1] 0.36881
[PC2] 0.13049	[PC2] 0.13049
[PC3] 0.11313	[PC3] 0.11313
[PC4] 0.10487	[PC4] 0.10487
[PC5] 0.07794	[PC5] 0.07794

Berdasarkan data hasil percobaan tersebut, dapat terlihat bahwa hasil dari *explained variance ratio* tidak terdapat perbedaan antara model buatan sendiri dan model *library*. Namun, model buatan sendiri memiliki waktu eksekusi yang lebih cepat daripada model *library*. Hal tersebut dapat dikarenakan pada implementasi model buatan sendiri, saya menggunakan *library* dari *numpy* untuk menghitung kovariansi matriks, *eigenvector*, serta *eigenvalue*.