

# Principal Component Analysis (PCA)

**Tim Asprak Metkuan**

<http://bit.ly/V2uKyl>

## Introduction

---

- Teknik untuk mengubah dimensi data menjadi data yang berdimensi lebih kecil, namun tetap mempertahankan sebagian besar (sekian persen cirinya).
- Suatu cara mengidentifikasi pola dalam suatu data.
- Baik untuk menganalisis data yang berdimensi besar.
- Tujuan PCA adalah untuk mendapatkan sesuatu yang berarti yang ada pada tabel atau matriks (memisahkan komponen utama dengan noise).

## Example

```
dataku =  
[ 7 26 6 60;  
 1 29 15 52;  
11 56 8 20;  
11 31 8 47;  
 7 52 6 33;  
11 55 9 22;  
 3 71 17 6;  
 1 31 22 44;  
 2 54 18 22;  
21 47 4 26;  
 1 40 23 34;  
11 66 9 12;  
10 68 8 12]
```

- **Four column of data**

## Get the covariance

```
covnya = cov(dataku)
```

Result :

```
covnya =
```

```
34.6026 20.9231 -31.0513 -24.1667  
20.9231 242.1410 -13.8782 -253.4167  
-31.0513 -13.8782 41.0256 3.1667  
-24.1667 -253.4167 3.1667 280.1667
```

## Using pcacov

```
[pc,eigenval,explained] = pcacov(covnya)
```

Result :

```
pc = //principal component dari data -> vector eigen
```

```
-0.0678  0.6460  -0.5673  0.5062
-0.6785  0.0200  0.5440  0.4933
0.0290  -0.7553  -0.4036  0.5156
0.7309  0.1085  0.4684  0.4844
```

```
eigenval = //menunjukkan tingkat kepentingan suatu kolom pada pc
```

```
517.7969
67.4964
12.4054
0.2372
```

```
explained = ->prosentase eigenval/total eigen val
```

```
86.5974
11.2882
2.0747
0.0397
```

## To get 97% information from data

- First row of explained = 86.5974%
- Second row of explained = 11.2882%
- Total = **97.8856%**
- So, take first and second column of pc

```
pc97=pc(:,[1:2])
```

- Result :

```
pc97 =
```

```
-0.0678  0.6460
-0.6785  0.0200
0.0290  -0.7553
0.7309  0.1085
```

To build new data (with 97% information) from PCA

```
datakubaru=(pc97' * dataku)'
```

- Result :  
datakubaru =  
25.9105 7.0189  
18.6960 -4.4628  
-23.8931 4.3530  
12.8034 6.7821  
-11.4645 4.6098  
-21.7238 3.7946  
-43.4995 -8.8318  
11.6951 -10.5779  
-20.1739 -8.8373  
-14.1953 14.3053  
-1.6913 -12.2380  
-36.4962 2.9297  
-37.8145 3.0790
- Only two column

## PCA for image

---

1. **>> datalmage = imread('nama\_file.ekstensi');**  
% Imread digunakan untuk membaca piksel citra dan menyimpannya dalam bentuk matriks
2. **>> whos datalmage**  
% melihat class citra
3. **>> figure, imshow(datalmage)**  
% menampilkan gambar
4. **>> datalmageD = double(datalmage);**  
% mengkonversi class uint8 ke double agar dapat diproses

## PCA for image (cont..)

---

```
5. >> gambar = rgb2gray(dataImage);  
    % mengubah citra menjadi grayscale  
6. >> gambarD = double(gambar);  
7. >> covnya=cov(gambarD);  
8. >> [pc,eigenval,explained] = pcacov(covnya)  
9. >> pcpersen=pc(:,[1:10])  
    %mengambil informasi dari kolom 1-10  
10. >> gambarbaru=(pcpersen'*gambarD)'  
11. >> figure, imshow(gambarbaru)
```

## Exercise

- Lakukan PCA pada data penyakit euthyroid (24 fitur tanpa kelas). Lakukan pengambilan informasi sebanyak:
  - 90%
  - 80%
- Berapa jumlah fitur akhir yang diperoleh untuk masing-masing persentase PCA?

## Exercise (cont..)

- Lakukan reduksi dimensi dengan PCA pada citra iris (setosa, virginica, dan versicolor).
  - gunakan persentase pca 90% (dimensi akhir bisa berbeda tiap citra)
  - matrix tiap proses dilampirkan (**pc, eigen, explained, pcpersen**)
  - citra hasil pca diprint screen
- Tentukan dimensi citra akhir yang diperoleh untuk masing-masing citra!

Subjek: NRP\_LKP 12

Kirim ke:

P1: deanimations@gmail.com

P2: haryanto.toto@gmail.com

**Terima Kasih**