



Laporan Tugas 2

Analisis Komponen Utama

Analisis Multivariat Kelas B
Kelompok L

Evan Haryowidyatna	2006485011
Muhammad Jauhar Hakim	2006463982
Siskawati Simandalahi	2006572970

Fakultas Matematika dan Ilmu Pengetahuan Alam
Universitas Indonesia
Depok
Februari 2022

I. Penjelasan Data

Data yang kami peroleh merupakan data kebakaran hutan yang terjadi pada tahun 2008. Kami mendapatkan data dari situs <https://archive.ics.uci.edu/ml/datasets/Forest+Fires>.

Data yang kami gunakan terdiri dari banyak variabel di antaranya adalah bulan, hari, beberapa satuan indeks seperti (FFMC, DMC, DC, ISI), temperatur udara, kelembaban udara, angin, hujan, dan luas daerah yang terbakar.

Kami memutuskan untuk hanya mengambil empat variabel pada analisa kali ini yaitu temperatur udara, kelembaban udara, angin, dan luas daerah yang terbakar dikarenakan variabel indeks terlihat kurang valid di mana variabel indeks merupakan data yang sudah diatur oleh campur tangan manusia sehingga menurut kami hasilnya kurang mencerminkan faktor alam yang sebenarnya.

Lalu kami tidak menggunakan variabel hujan pada analisa kami karena pada data yang didapatkan, data pada variabel hujan tidak memberikan informasi yang jelas di mana datanya kebanyakan berisi angka 0. Data yang kami analisa bisa dilihat di gambar berikut. Dan berikut ini merupakan link data set yang dipakai untuk analisis komponen utama kali ini : <https://bit.ly/datasetforestfires>.

temp	RH	wind	area
18.0	42	2.7	0.36
21.7	38	2.2	0.43
21.9	39	1.8	0.47
23.3	31	4.5	0.55
21.2	51	8.9	0.61
16.6	53	5.4	0.71
23.8	32	5.4	0.77
27.4	22	4.0	0.90
13.2	40	5.4	0.95
24.2	28	3.6	0.96
17.4	43	6.7	1.07
23.7	25	4.5	1.12
23.2	39	5.4	1.19
24.8	29	2.2	1.36
24.6	43	1.8	1.43

temp	RH	wind	area
20.1	47	4.9	1.46
29.6	27	2.7	1.46
16.4	47	1.3	1.56
28.6	27	2.2	1.61
18.4	45	3.6	1.63
20.5	35	4.0	1.64
19.0	34	5.8	1.69
16.1	29	3.1	1.75
20.3	41	4.0	1.90
15.2	31	8.5	1.94
17.8	56	1.8	1.95
17.8	67	2.2	2.01
5.3	70	4.5	2.14
16.6	47	0.9	2.29
23.4	33	4.5	2.51

II. Tujuan

Tujuan dari Analisis Komponen Utama adalah menyederhanakan suatu data, dengan cara mentransformasi linear sehingga terbentuk sistem koordinat baru dengan variansi maksimum. Analisis Komponen Utama digunakan untuk mereduksi dimensi suatu data tanpa mengurangi karakteristik data tersebut secara signifikan.

III. Dasar Teori

Analisis Komponen Utama adalah teknik yang digunakan untuk menyederhanakan suatu data, dengan cara mentransformasi linear sehingga terbentuk sistem koordinat baru dengan variansi maksimum. Metode ini mengubah dari sebagian besar variabel asli yang saling berkorelasi menjadi satu himpunan variabel baru yang lebih kecil dan saling bebas (tidak berkorelasi lagi). Prinsip dasar dari algoritma Principal Component Analysis adalah mengurangi satu set data namun tetap mempertahankan sebanyak mungkin variasi dalam set data tersebut.

Dalam melakukan proses analisis komponen utama, terdapat dua tipe langkah di mana yang satu menggunakan matriks korelasi dan yang lainnya menggunakan matriks kovariansi. Kedua tipe ini dipilih berdasarkan proporsi variansi pada data dengan melihat mana yang lebih relevan.

Terdapat beberapa kriteria dalam menentukan variabel atau faktor mana yang akan dipertahankan di antaranya adalah,

1. Pertahankan komponen yang cukup untuk memperhitungkan persentase tertentu dari total varians, katakanlah 80%.
2. Pertahankan komponen yang nilai eigen nya lebih besar dari rata-rata nilai eigen, $\sum_{i=1}^p \frac{\lambda_i}{p}$. Untuk matriks korelasi, rata-rata ini adalah 1.
3. Gunakan grafik scree plot untuk melihat selisih nilai eigen "besar" dan nilai eigen "kecil".
4. Uji signifikansi komponen "lebih besar", yaitu komponen yang sesuai dengan nilai eigen yang lebih besar.

IV. Pemodelan

Berikut ini adalah langkah kerja untuk melakukan analisis komponen utama :

1. Mengimpor data kebakaran hutan
2. Mencari vektor mean, matriks kovariansi, dan matriks korelasi dari data kebakaran hutan
3. Mencari nilai eigen dan vektor eigen dari matriks kovariansi dan matriks korelasi
4. Menentukan berapa banyak komponen yang harus dipertahankan (akan digunakan metode 1 dan metode 3)
5. Membandingkan analisis komponen utama yang berasal dari matriks kovariansi dan analisis komponen utama yang berasal dari matriks korelasi
6. Transformasi linier kan data awal ke data ke dalam dimensi baru yang telah direduksi oleh teknik analisis komponen utama.
7. Visualisasikan beberapa plot yang dibutuhkan untuk menjelaskan analisis komponen utama

V. Proses komputasi

1. Import File

```
data <- read.table("forestfiresfinal.csv", header=TRUE, sep=";")  
datanew <- as.matrix(data)  
datanew
```

2. Vektor Mean, Matriks Kovariansi, Matriks Korelasi

#Vektor Mean

```
matrix(colMeans(datanew), ncol=1)
```

#Matriks Kovariansi

```
covdata <- cov(datanew)
```

```
cov(datanew)
```

```
#Matriks Korelasi
```

```
cordata <- cor(datanew)
```

```
cordata
```

3. Nilai Eigen dan Vektor Eigen dari Matriks Kovariansi

```
#Nilai Eigen dari Matriks Kovariansi
```

```
eig_val_cov <- eigen(covdata)$values
```

```
eig_val_cov
```

```
#Vektor Eigen dari Matriks Kovariansi
```

```
eig_vec_cov <- eigen(covdata)$vectors * (-1)
```

```
eig_vec_cov
```

4. Nilai Eigen dan Vektor Eigen dari Matriks Korelasi

```
#Nilai Eigen dari Matriks Korelasi
```

```
eig_val_cor <- eigen(cordata)$values
```

```
eig_val_cor
```

```
#Vektor Eigen dari Matriks Korelasi
```

```
eig_vec_cor <- eigen(cordata)$vectors * (-1)
```

```
eig_vec_cor
```

5. Menentukan berapa banyak komponen yang harus dipertahankan

Metode 1 :

Menghitung proporsi variansi dan kumulatifnya - Matriks Kovariansi

```
pvcov <- eig_val_cov/sum(eig_val_cov)

kumvcov <- cumsum(pvcov)

data.pr<-prcomp(data, scale. = FALSE)

summary(data.pr)

pvcov

Kumvcov
```

Menghitung proporsi variansi dan kumulatifnya - Matriks Korelasi

```
pvcor <- eig_val_cor/sum(eig_val_cor)

kumvcor <- cumsum(pvcor)

data.pr<-prcomp(data, scale. = TRUE)

summary(data.pr)

pvcor

Kumvcor
```

Metode 3

Scree Graph – PCA dari Matriks Kovariansi

```
data.pr<-prcomp(data, scale. = FALSE)
screeplot(data.pr, type = "l", npcs = 4, main = "Screeplot from Covariance
Matrix")
abline(h = 1, col="red", lty=5)
legend("topright", legend=c("Eigenvalue = 1"),col=c("red"), lty=5, cex=0.6)
```

Scree Graph – PCA dari Matriks Korelasi

```
data.pr<-prcomp(data, scale. = TRUE)
screeplot(data.pr, type = "l", npcs = 4, main = "Screeplot from Corellation
Matrix")
abline(h = 1, col="red", lty=5)
legend("topright", legend=c("Eigenvalue = 1"),col=c("red"), lty=5, cex=0.6)
```

6. Transformasi linier data analisis komponen utama

PCA dari Matriks Kovariansi

#pca versi cov

```
pcajadi_cov <- scale(datanew) %*% (eig_vec_cov * (-1))
```

```
head(pcajadi_cov) #6 data pertama
```

PCA dari Matriks Korelasi

#pca versi cor

```
pcajadi_cor <- scale(datanew) %*% (eig_vec_cor * (-1))
```

```
head(pcajadi_cor) #6 data pertama
```

PCA dengan package R dan Visualisasi PCA

```
data.pr<-prcomp(data, scale. = TRUE)
```

```
cumpro <- cumsum(data.pr$sdev^2 / sum(data.pr$sdev^2))
```

```
plot(cumpro[0:15], xlab = "PC #", ylab = "Amount of explained variance",  
main = "Cumulative variance plot")
```

```
abline(v = 6, col="blue", lty=5)
```

```
abline(h = 0.88759, col="blue", lty=5)
```

```
legend("topleft", legend=c("Cut-off @ PC6"),col=c("blue"), lty=5, cex=0.6)
```

```
plot(data.pr$x[,1],data.pr$x[,2], xlab="PC1 ", ylab = "PC2 ", main = "PC1  
PC2 - plot")
```

```
data.pr
```

```
data.pr$x[,1:3]
```

VI. Hasil Komputasi

1. Import File

temp	RH	wind	area
18.0	42	2.7	0.36
21.7	38	2.2	0.43
21.9	39	1.8	0.47
23.3	31	4.5	0.55
21.2	51	8.9	0.61
16.6	53	5.4	0.71
23.8	32	5.4	0.77
27.4	22	4.0	0.90
13.2	40	5.4	0.95
24.2	28	3.6	0.96
17.4	43	6.7	1.07
23.7	25	4.5	1.12
23.2	39	5.4	1.19
24.8	29	2.2	1.36
24.6	43	1.8	1.43

temp	RH	wind	area
20.1	47	4.9	1.46
29.6	27	2.7	1.46
16.4	47	1.3	1.56
28.6	27	2.2	1.61
18.4	45	3.6	1.63
20.5	35	4.0	1.64
19.0	34	5.8	1.69
16.1	29	3.1	1.75
20.3	41	4.0	1.90
15.2	31	8.5	1.94
17.8	56	1.8	1.95
17.8	67	2.2	2.01
5.3	70	4.5	2.14
16.6	47	0.9	2.29
23.4	33	4.5	2.51

2. Vektor Mean, Matriks Kovariansi, Matriks Korelasi

2.1. Vektor Mean

20.336667
39.700000
3.950000
1.347333

2.2. Matriks Kovariansi

	temp	RH	wind	area
temp	24.4513678	-39.616207	-1.7381034	-0.8004161
RH	-39.6162069	137.320690	-1.1500000	1.6391724
wind	-1.7381034	-1.1500000	3.9053448	-0.1821724
area	-0.8004161	1.639172	-0.1821724	0.3477857

2.3. Matriks Korelasi

	temp	RH	wind	area
temp	1.0000000	-0.68368071	-0.17786658	-0.2744785
RH	-0.6836807	1.00000000	-0.04965925	0.2371924
wind	-0.1778666	-0.04965925	1.00000000	-0.1563137
area	-0.2744785	0.23719239	-0.15631370	1.0000000

3. Nilai Eigen dan Vektor Eigen dari Matriks Kovariansi

3.1. Nilai Eigen

$$\lambda_1 = 149.861, \quad \lambda_2 = 12.4107, \quad \lambda_3 = 3.44815, \quad \lambda_4 = 0.30467$$

3.2. Vektor Eigen

-0.301217177	-0.92827974	0.21497894	-0.0367286072
0.953471020	-0.29254577	0.07286593	0.0007378772
-0.003940523	0.22885730	0.97107848	-0.0679367231
0.012070592	0.01838566	-0.07403514	-0.9970130724

4. Nilai Eigen dan Vektor Eigen dari Matriks Korelasi

4.1. Nilai Eigen

$$\lambda_1 = 1.840, \quad \lambda_2 = 1.135, \quad \lambda_3 = 0.751, \quad \lambda_4 = 0.272$$

4.2. Vektor Eigen

-0.65734203	-0.23091565	-0.08911436	-0.7117851
0.64353988	0.01744482	0.40308512	-0.6504418
0.02835677	0.84714401	-0.47226295	-0.2418897
0.39109305	-0.47824750	-0.77881238	-0.1085211

5. Menentukan berapa banyak komponen yang harus dipertahankan

5.1. Proporsi variansi dan kumulatifnya - Matriks Kovariansi

Importance of components:

	PC1	PC2	PC3	PC4
Standard deviation	12.2418	3.52289	1.85692	0.55197
Proportion of Variance	0.9026	0.07475	0.02077	0.00184
Cumulative Proportion	0.9026	0.97740	0.99816	1.00000

Satu komponen utama pertama menjelaskan 90.26% dari total variansi.

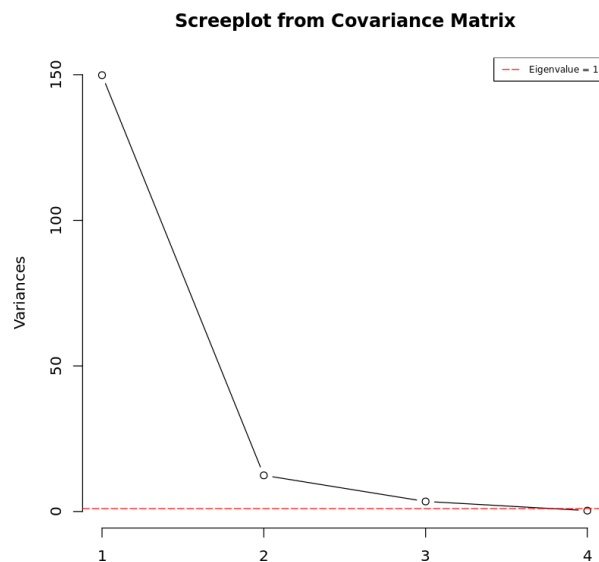
5.2. Proporsi variansi dan kumulatifnya - Matriks Korelasi

Importance of components:

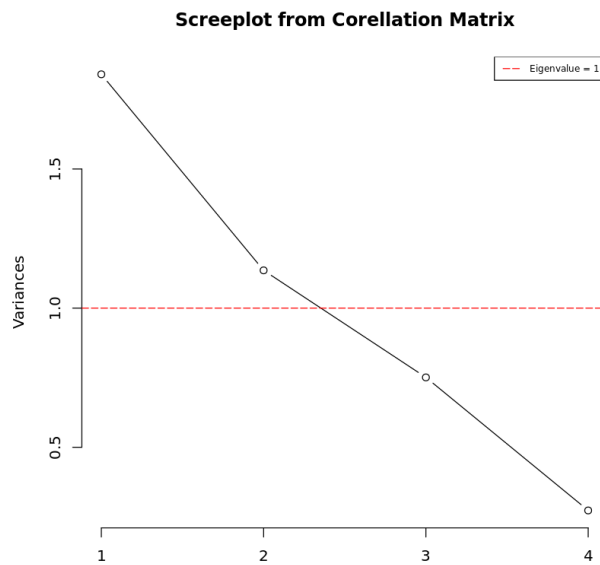
	PC1	PC2	PC3	PC4
Standard deviation	1.3566	1.0657	0.8666	0.52244
Proportion of Variance	0.4601	0.2839	0.1878	0.06824
Cumulative Proportion	0.4601	0.7440	0.9318	1.00000

Tiga komponen utama pertama menjelaskan 93.18% dari total variansi.

5.3. Scree Graph – PCA dari Matriks Kovariansi



5.4. Scree Graph – PCA dari Matriks Korelasi



5.5 Kesimpulan

Berdasarkan 5.1 – 5.4 kami memilih melakukan analisis komponen utama dari matriks korelasi karena variansi untuk variabel pertama sangat besar yang menyebabkan bias. Sehingga banyaknya komponen yang harus dipertahankan sebanyak tiga komponen pertama yaitu PC1, PC2, dan PC3.

6. Transformasi linier data analisis komponen utama

6.1. PCA dari Matriks Kovariansi (6 data pertama)

Berikut ini 6 data pertama yang sudah diubah ke dimensi yang baru dengan metode analisis komponen utama dari matriks kovariansi (data sudah di standarisasi).

-0.3117635	-0.2056971	0.5775714	-1.729674
0.2366557	0.4447562	0.6960664	-1.600786
0.1658571	0.5523421	0.8827300	-1.545489
0.9058081	0.3002697	-0.4450957	-1.306518
-0.8418761	-0.1060858	-2.6327341	-1.070678
-1.2938441	-0.5174946	-0.7127707	-1.056231

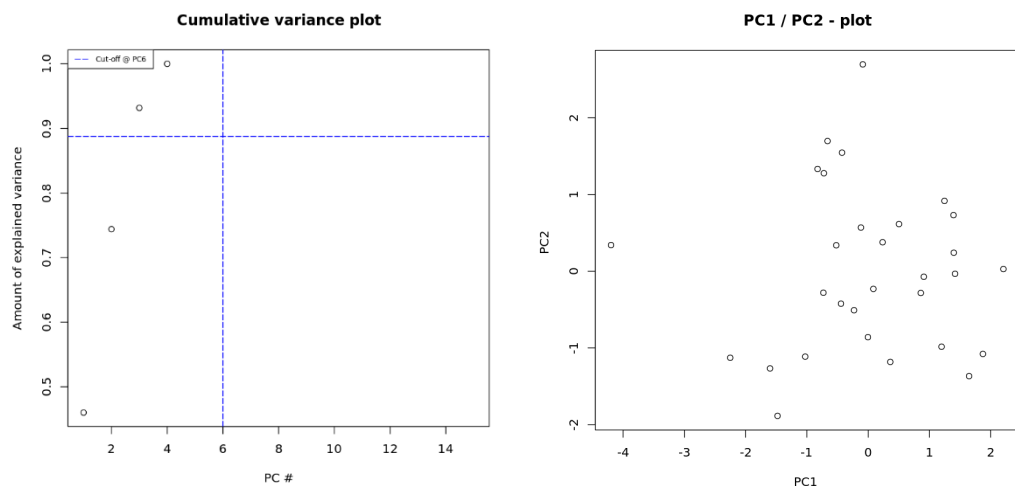
6.2. PCA dari Matriks Korelasi (6 data pertama)

Berikut ini 6 data pertama yang sudah diubah ke dimensi yang baru dengan metode analisis komponen utama dari matriks korelasi (data sudah di standarisasi).

0.23577090	-0.37738250	-1.7238347	-0.5433766
0.90805204	0.07246024	-1.5466085	-0.2811231
0.85893490	0.28421921	-1.6201675	-0.2384280
1.39258448	-0.73103818	-0.5688715	-0.1357470
-0.08784757	-2.69638824	-0.1639389	1.2216981
-0.82527818	-1.33272220	-1.0199882	0.2605576

6.3. PCA dengan package R dan Visualisasi PCA

Berikut ini merupakan plot PC1/PC2 dan plot dari proporsi kumulatif variansinya dari metode analisis komponen utama dari matriks korelasi.



VII. Pembahasan

Berdasarkan proses komputasi yang telah dilakukan sebelumnya, diperoleh vektor mean, matriks kovariansi **S** dan matriks korelasi **R**. Di sini kami mencari matriks kovariansi dan korelasi dengan tujuan agar bisa mencoba kedua metode saat menentukan proporsi variansi dan kumulatifnya.

Sebelum itu kami mencari nilai eigen dan vektor eigen dari kedua matriks tersebut yaitu matriks **S** dan **R**. Lalu kami mencari proporsi variansi dan kumulatifnya dengan menggunakan kedua matriks tersebut.

Saat menggunakan matriks kovariansi kami melihat bahwa Satu komponen utama pertama menjelaskan 90.26% dari total variansi, dan saat kami menggunakan matriks korelasi kami melihat bahwa Tiga komponen utama pertama menjelaskan 93.17% dari total variansi.

Hal ini disebabkan karena variansi variabel pertama sangat lah besar sehingga membuat kami berpendapat bahwa metode analisis komponen utama dari matriks kovariansi memberikan hasil yang bias sebagaimana dapat dilihat bahwa hanya dengan satu komponen utama pertama, sudah dapat menjelaskan lebih dari 90% dari total variansi data. Sehingga kami melakukan analisis komponen utama yang berasal dari matriks korelasi.

Berdasarkan metode 1 (metode proporsi kumulatif) yaitu tiga komponen pertama menjelaskan 93.17% dari total variansi dan metode 3 (metode scree graph) terlihat patahan besar setelah komponen ke 3 kami mempertahankan tiga komponen pertama yaitu PC1, PC2, dan PC3. Lalu kami melakukan transformasi linier dari data yang lama ke dalam data yang baru dengan koordinat yang baru yaitu PC1, PC2, dan PC3.

Berikut ini merupakan data yang sudah ditransformasi linier dengan metode analisis komponen utama dari matriks korelasi:

PC1	PC2	PC3	PC1	PC2	PC3
0.235770896	0.37738250	-1.72383467	-0.520704802	0.33779232	0.12044870
0.908052044	-0.07246024	-1.54660849	1.872089002	-1.07869936	0.45386093
0.858934902	-0.28421921	-1.62016748	-1.027224690	-1.11374705	-0.67448331
1.392584480	0.73103818	-0.56887147	1.646853098	-1.36798145	0.51444382
-0.087847574	2.69638824	-0.16393890	-0.730945172	-0.28093688	0.07244329
-0.825278183	1.33272220	-1.01998815	0.085018127	-0.23052996	0.56306194
1.245323528	0.91657449	-0.08864412	-0.118454485	0.56909510	1.06661457
2.206937798	0.02800531	0.15732448	-0.230428334	-0.50900115	0.62034184
-0.722496131	1.27751473	-0.31714440	-0.443495350	-0.42310660	0.69643218
1.417992349	-0.03375493	-0.12308425	-0.663391826	1.69676502	2.07671435
-0.427153983	1.54580967	0.12449664	-1.601179745	-1.26766203	-0.32430307
1.397254730	0.24118239	0.39747494	-2.250797495	-1.12847420	-0.52784981
0.502611871	0.61441307	0.21441854	-4.196453177	0.34024847	-0.13498929
1.197657614	-0.98481150	0.04701170	-1.479011345	-1.88655372	0.19758124
0.361549235	-1.18286705	-0.44130691	-0.003767382	-0.86012635	1.95254517