STA 437/2005: Methods for Multivariate Data

Week 7: Principal Component Analysis

Piotr Zwiernik

University of Toronto

Example 1: Decathlon

The columns are 10 different disciplines in decathlon:

```
> data("olympic", package = "ade4")
> athletes = setNames(olympic$tab,
+   c("m100", "long", "weight", "high", "m400", "m110", "disc", "pole", "javel", "m1500"))
> head(athletes)
   m100 long weight high  m400  m110  disc pole javel  m1500
1 11.25 7.43  15.48 2.27 48.90 15.13 49.28  4.7 61.32 268.95
2 10.87 7.45  14.97 1.97 47.71 14.46 44.36  5.1 61.76 273.02
3 11.18 7.44  14.20 1.97 48.29 14.81 43.66  5.2 64.16 263.20
4 10.62 7.38  15.02 2.03 49.06 14.72 44.80  4.9 64.04 285.11
5 11.02 7.43  12.92 1.97 47.44 14.40 41.20  5.2 57.46 256.64
6 10.83 7.72  13.58 2.12 48.34 14.18 43.06  4.9 52.18 274.07
```

PCA Biplot for Decathlon data

PCA Biplot of Olympic Athletes

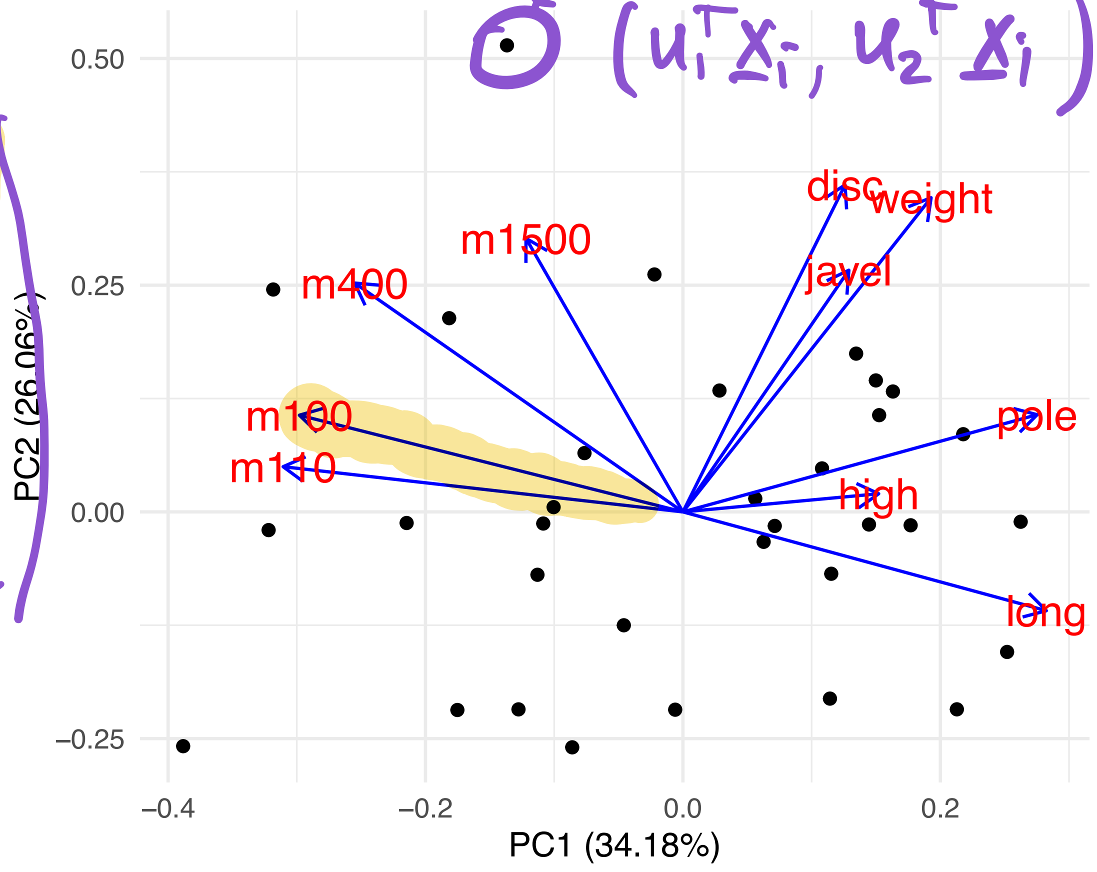
loadings

$$U_q = \begin{bmatrix} u_{11} & u_{12} \\ u_{21} & u_{22} \\ \vdots & \vdots \\ u_{m1} & u_{m2} \end{bmatrix}$$

$$\begin{array}{cc} \uparrow & \uparrow \\ u_1 & u_2 \end{array}$$

proj. of $\underline{x}_i \in \mathbb{R}^{10}$

$(u_1^T \underline{x}_i, u_2^T \underline{x}_i)$



Example 2: Pottery

Chemical analysis data on Romano-British pottery made in three different regions (kiln 1, kilns 2-3, and kilns 4-5):

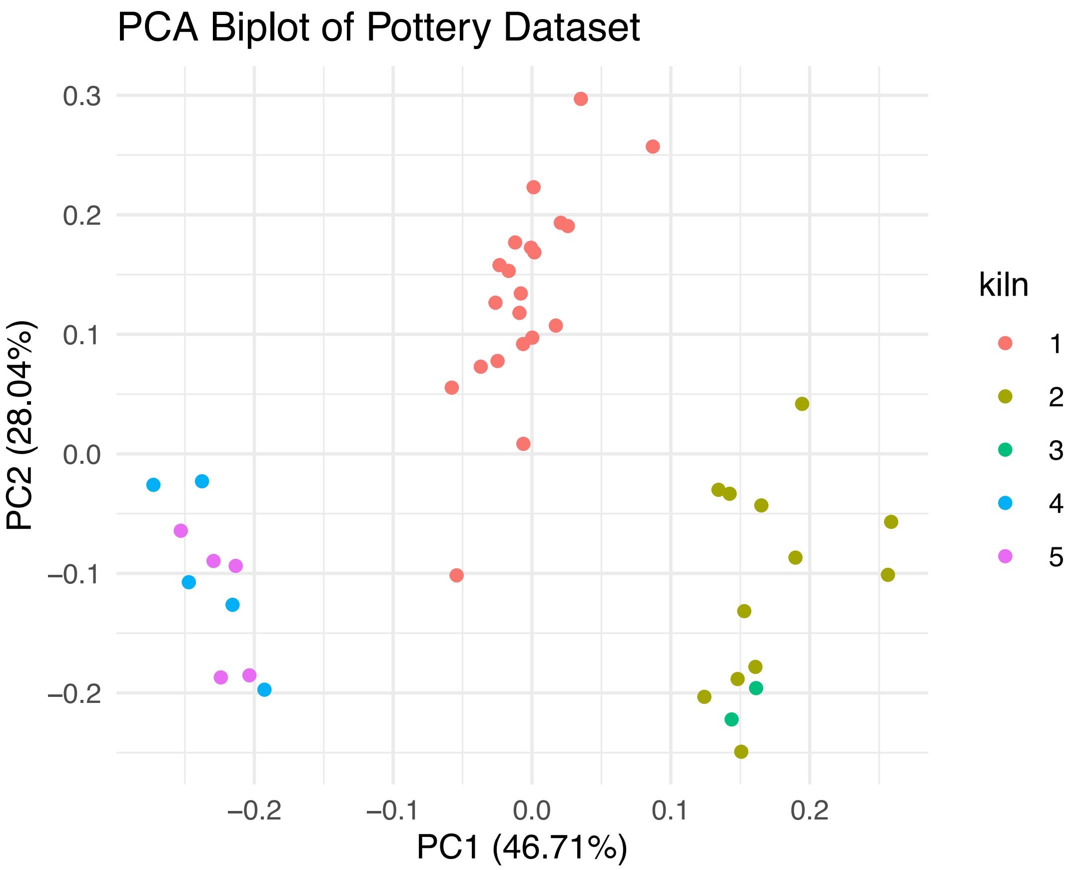
```
> data("pottery", package = "HSAUR2")
> head(pottery)
```

```
  Al2O3 Fe2O3  MgO  CaO Na2O  K2O TiO2   MnO   BaO kiln
1  18.8  9.52 2.00 0.79 0.40 3.20 1.01 0.077 0.015    1
2  16.9  7.33 1.65 0.84 0.40 3.05 0.99 0.067 0.018    1
3  18.2  7.64 1.82 0.77 0.40 3.07 0.98 0.087 0.014    1
4  16.9  7.29 1.56 0.76 0.40 3.05 1.00 0.063 0.019    1
5  17.8  7.24 1.83 0.92 0.43 3.12 0.93 0.061 0.019    1
6  18.8  7.45 2.06 0.87 0.25 3.26 0.98 0.072 0.017    1
```

Question: Do the chemical profiles of each pot suggest different types of pots and if any such types are related to kiln or region.

PCA Biplot for Pottery data

PCA Biplot of Olympic Athletes



Lecture 7 : PCA.

$X = (X_1, \dots, X_m) \sim (\mu, \Sigma)$ $\quad \sum u_i X_i$ $\quad \parallel$

Problem: find $u \in \mathbb{R}^m$ s.t $Var(u^T X)$ max.
$\|u\| = 1$

Recall : $Var(u^T X) = u^T \Sigma u$

maximize $f(u) = u^T \Sigma u$ $\quad$ s.t $\quad u^T u = 1$

the Lagrangian :

$$\mathcal{L} = u^T \Sigma u - \lambda (u^T u - 1)$$

$$\nabla \mathcal{L} = 2 \Sigma u - 2 \lambda u = 0$$

stat. points $\quad \underline{\Sigma u = \lambda u}$ $\quad$ u is an eigenvector of Σ with eigenvalue λ.

there is m orth. eigenvectors $u_1, \dots, u_m$ with eigenv. $\lambda_1, \dots, \lambda_m$

at u_i, $f(u_i) = u_i^T \underbrace{\Sigma u_i}_{\lambda_i \cdot u_i} = \lambda_i \underbrace{u_i^T u_i}_{1} = \lambda_i$

if $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m$

then the maximum is u_1.

$Var(u_1^T X)$ is the largest, $Z_1 = u_1^T X$

Problem 2 : Find u, $\|u\| = 1$ st.

$u^T X$ uncorrelated with $u_1^T X$

$u^T X$ uncorrelated with $u_1^T X$
and has the largest variance.

$$\underset{=0}{\mathrm{Cov}}(u^T X, u_1^T X) = u^T \mathrm{Cov}(X,X) u_1$$

$$= u^T \underbrace{\Sigma u_1}_{\lambda_1 u_1} = \underbrace{\lambda_1}_{>0} u^T u_1$$

In other words $u \perp u_1$

Maximize $u^T \Sigma u$ s.t $u^T u = 1$ and $u^T u_1 = 0$

$$\mathcal{L} = u^T \Sigma u - \lambda (u^T u - 1) - \nu u_1^T u$$

$$\nabla \mathcal{L} = 2\Sigma u - 2\lambda u - \nu u_1$$

$$\underbrace{\Sigma u - \lambda u = \frac{\nu}{2} u_1}$$ $= 0$

Claim $\nu = 0$ if not

$$\underbrace{u_1^T \Sigma}_{\lambda_1 u_1^T} u - \lambda u_1^T u = \frac{\nu}{2} u_1^T u_1$$

$$(\lambda_1 - \lambda) \underbrace{u_1^T u}_{= 0} = \frac{\nu}{2} \underbrace{u_1^T u_1}_{1} = \frac{\nu}{2}$$

so $\Sigma u = \lambda u$ again . . .

the solution is u_2 with the second largest eigenvalue

$$Z_2 = u_2^T X$$

$$\vdots$$

$$Z_m = u_m^T X$$

u_i's called canonical directions

$\lambda_1 \geq \dots \geq \lambda_m$

Recall : $\Sigma = U \Lambda U^T$

$$U = \begin{bmatrix} | & & | \\ u_1 & \cdots & u_m \\ | & & | \end{bmatrix} \in O(m)$$

<u>Check :</u>

$$\Sigma u_i = U \Lambda \underbrace{U^T u_i}_{= e_i} = \lambda_i U \cdot e_i = \lambda_i u_i$$

e_i ith canonical vec.

$(0, \dots, \underset{i}{1}, 0 \dots, 0)$

$\vdots$

DATA $\underline{x}_1, \dots, \underline{x}_n \in \mathbb{R}^m$

$\underline{X} \in \mathbb{R}^{n \times m} \to$ standardized

(i) $HX = X$

$\left(\mathbb{1}^T X = 0^T \right)$

(ii) $\mathrm{diag}(S_n) = (1, \dots, 1)$

$$\frac{1}{n} \sum_{i=1}^{n} (\underline{x}_i)_k^2 = 1$$

sample <u>correlation</u> S_n

$$S_n = U \Lambda U^T$$

columns of U are the principal directions.

Say we take the first d principal directions $u_1, \ldots, u_d$

$\underline{x}_1, \ldots, \underline{x}_n \in \mathbb{R}^m$

define SCORES

$y_1, \ldots, y_n \in \mathbb{R}^d$

projections of $\underline{x}_i$'s onto the linear subspace spanned by $u_1, \ldots, u_d$

$$y_i = \left(u_1^T \underline{x}_i, \ldots, u_d^T \underline{x}_i \right)$$

If $d = 2$ biplot

$d \ll m$ without removing important inf.

$\underline{\lambda_i = u_i^T S_n u_i}$ the corresp. eingenval.

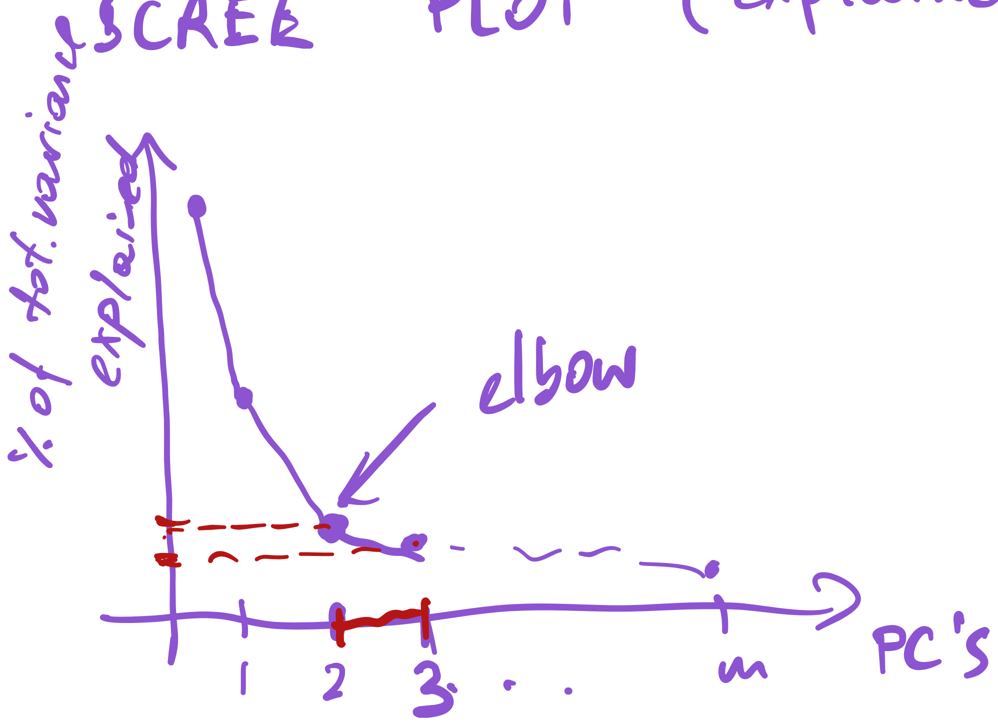
$$\mathrm{tr}(S_n) = \mathrm{tr}(U \Lambda U^T) = \lambda_1 + \dots + \lambda_m$$

$$\sum_{i=1}^{m} (S_n)_{ii} \; (= m)$$

$\hookrightarrow$ total variance (sum of all variances)

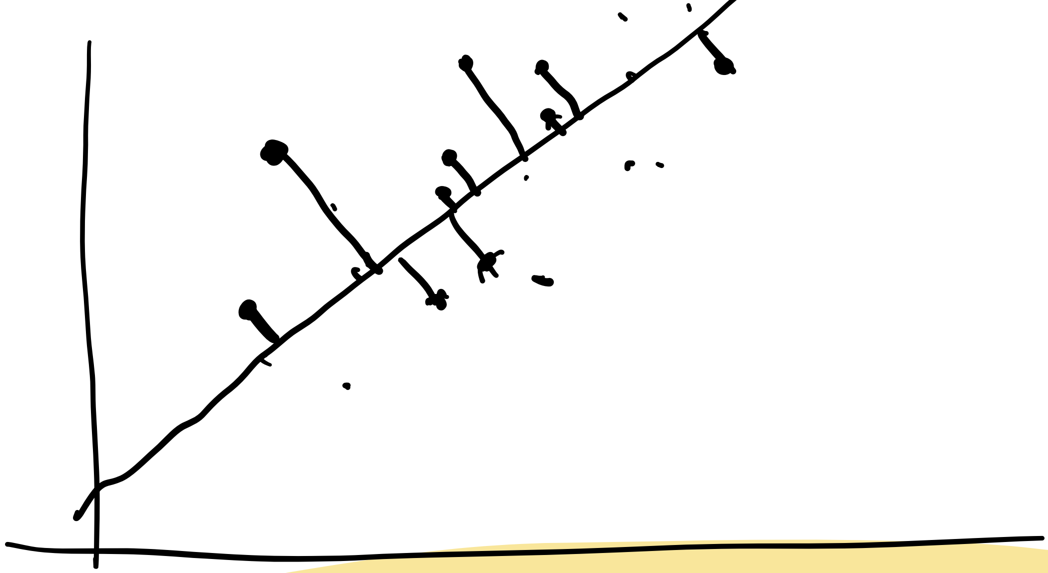
$$\lambda_i = \text{variance of } (u_i^T \underline{x}_1, \dots, u_i^T \underline{x}_n)$$

SCREE PLOT (explained variance)



$$\frac{\lambda_i}{m} \cdot 100\%.$$

PCA and AFFINE SUBSPACE APPROX.



Assume data centred

Approximate the data by a d-dim. affine subspace

$$\mu + \mathrm{span}(w_1, \dots, w_d)$$

lin. indep.

every point $\underline{x}$ in this subspace is of the form

$$x = \mu + Wz \qquad z \in \mathbb{R}^d$$

$$\hookrightarrow \begin{pmatrix} | & & | \\ w_1 & \cdots & w_d \\ | & & | \end{pmatrix}$$

$$\boxed{\begin{aligned} &\text{minimize} \sum_{i=1}^{n} \|\underline{x}_i - (\mu + W z_i)\|^2 \qquad (*) \\ &\text{s.t.} \quad \mu \in \mathbb{R}^m, \; W \in \mathbb{R}^{m \times d}, \; \underline{z}_i \in \mathbb{R}^d \end{aligned}}$$

$-Wz_i$

$$\|x_i - \mu - Wz_i\|^2 = \|x_i - \mu\|^2 + \|Wz_i\|^2 - 2(x_i - \mu)^T W z_i$$

1. optimize (*) wrt μ.

$$\nabla_\mu (*) = \sum_{i=1}^{n} \left(2(\mu - \underline{x}_i) + 2\,W z_i \right)$$

$$\nabla_\mu (*) = 0 \quad \text{if}$$

$$n \cdot \mu - \sum_{i=1}^{n} \underline{x}_i + W \cdot \sum_{i=1}^{n} \underline{z}_i = 0$$

$$\hat{\mu} = \underbrace{\bar{\underline{x}}_n}_{= 0} - W \cdot \bar{\underline{z}}_n = -W \bar{\underline{z}}$$

min. $\displaystyle\sum_{i=1}^{m} \|x_i - W(z_i - \bar{z})\|^2$

equiv.

$$(**)\ \text{minim.}\ \sum_{i=1} \|x_i - W z_i\|^2 \qquad z_1, \dots, z_n \in \mathbb{R}^d \quad W \in \mathbb{R}^{n\times d}$$

forget. where z_i's are centered

$$\sum_{i=1}^n z_i = 0$$

Recall: $M \in \mathbb{R}^{m\times n}$

$$\|M\|_F^2 = \operatorname{tr}(MM^T) = \operatorname{tr}(M^TM)$$

$$= \sum_{i=1}^m \sum_{j=1}^n M_{ij}^2$$

$$\boxed{\|\underbrace{\underline{X}}_{n\times m} - \underbrace{\underline{Z}}_{n\times d} \cdot W^T\|_F^2} =$$

$$\underline{Z} = \begin{pmatrix} - z_1 - \\ \\ - z_n - \end{pmatrix}$$

$$= \sum_{i=1}^n \boxed{\sum_{j=1}^m \left(X_{ij} - (ZW^T)_{ij}\right)^2}$$

$$= \sum_{i=1}^{n} \| \underline{x}_i - W z_i \|^2$$

① $(**) = \| X - Z W^T \|_F^2$

② $Z \in \mathbb{R}^{n \times d}$, $W \in \mathbb{R}^{m \times d}$ otherwise unrestricted

so the only restriction on ZW^T is that it has rank $\leq d$.

$(***)$ minimize $\| X - M \|_F^2$

s.t $\operatorname{rank}(M) \leq d$

Theorem (Eckart-Young)

$\underline{X} = V \underset{n \times m}{D} U^T \in \mathbb{R}^{n \times m}$

the solution to $(***)$

is $\hat{M} = V \begin{pmatrix} \sigma_1 & & & & & \\ & \ddots & & & & \\ & & \sigma_d & & & \\ & & & 0 & & \\ & & & & \ddots & \\ & & & & & 0 \\ & & 0 & & & \end{pmatrix} U^T$