Principal Components Analysis

$$\underset{n \times p}{X} \qquad \overbrace{\text{col Means}}^{\text{makes things easier}}(X) = 0$$

observations predictors

Covariance Matrix

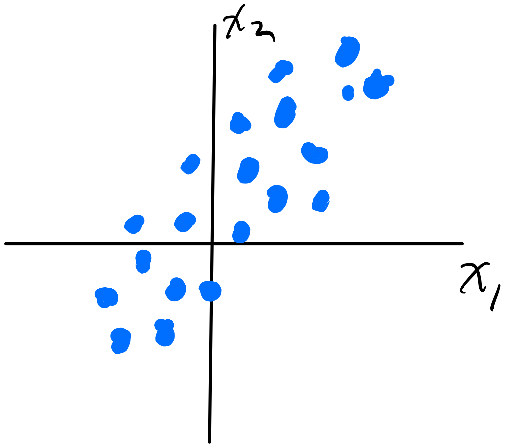
$$\underset{p \times p}{\text{cov}(X)} = \begin{bmatrix} Var(x_1) & & Cov(x_i, x_j) \\ & \ddots & \\ Cov(x_i, x_j) & & Var(x_p) \end{bmatrix}$$

In general $\text{Cov}(X) = \frac{1}{n-1}(X - \bar{X})^T (X - \bar{X})$

$n \times p$ matrix of p column means

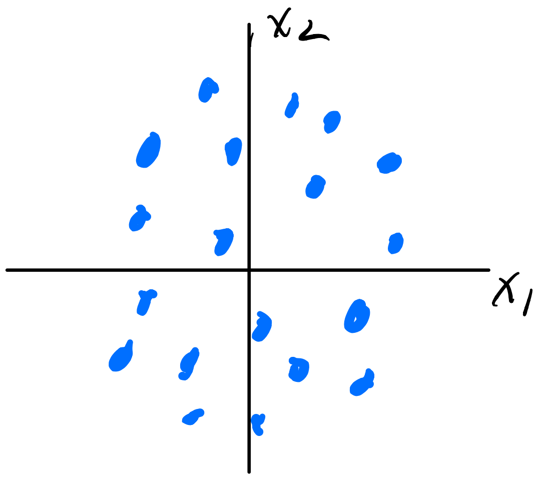
In our case $\text{Cov}(X) = \frac{1}{n-1} X^T X$

Examples



$cov(x_1, x_2) > 0$

x_1, x_2 correlated



$cov(x_1, x_2) \approx 0$

$x_1, x_2 \approx$ independent

SVD of $\underset{n\times p}{X} = \underset{n\times p}{U}\,\underset{p\times p}{\Sigma}\,\underset{p\times p}{V^T}$ colMeans(X) = 0

$\underset{p\times p}{U U^T} = \underset{p\times p}{V V^T} = I$ $\Sigma = \mathrm{diag}(\sigma_1, .. \sigma_p)$

$\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_p \geq 0$

Assume $\sigma_p > 0$
"full rank"

$\mathrm{Cov}(X) = \underset{p\times p}{X^T X}$

$$= (U\Sigma V^T)^T (U \Sigma V^T)$$
$$= V \Sigma \underbrace{U^T U}_{I} \Sigma V^T$$
$$= V \Sigma^2 V^T \quad \Sigma^2 = \mathrm{diag}(\sigma_1^2, \ldots, \sigma_p^2)$$

$V V^T = id$

$V = [V_1, \ldots, V_p]$ $V_1, V_2, \ldots, V_p$
orthnormal basis
for $\mathbb{R}^p$

Change of Basis

$V =$ Change of Basis Matrix

V- basis $\rightarrow$ Standard basis

$v_1, v_2 \dots v_p$ $\quad$ $e_1, e_2, \dots, e_p$

i.e. if $[x]_V = \begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_p \end{bmatrix}$ $\quad [x]_\mathcal{E} = V\begin{bmatrix} \tilde{x}_1 \\ \vdots \\ \tilde{x}_p \end{bmatrix}$

$V^{-1} = V^T$ Change of Basis Matrix

Standard basis $\rightarrow$ V- basis

$e_1, e_2, \dots, e_p$ $\quad$ $v_1, v_2 \dots v_p$

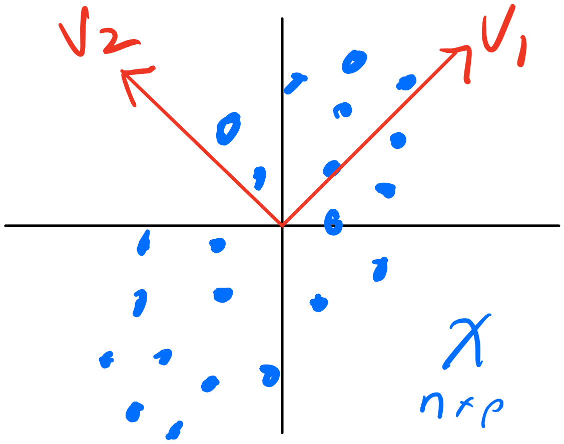
i.e. if $[x]_\mathcal{E} = \begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$ $\quad [x]_V = V^T\begin{bmatrix} x_1 \\ \vdots \\ x_p \end{bmatrix}$

Notice $\underset{n \times p}{X}$ each row is observation $1 \times p$

$\underset{p \times n}{X^T}$ each column is observation $p \times 1$

Change $\underset{n \times p}{X}$ (Standard Basis) $\rightarrow$ $\underset{n \times p}{\tilde{X}}$ (V basis)

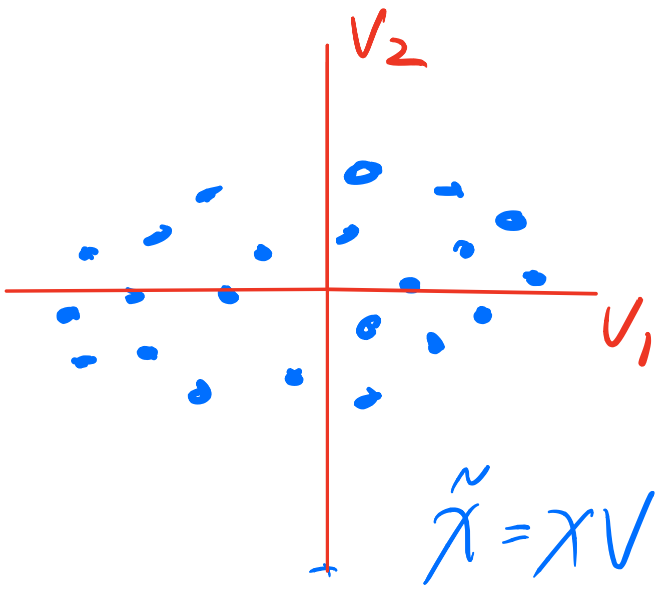
$$\underbrace{\underset{n \times p}{X}}_{\text{Data in Standard Basis}} \rightarrow \underset{p \times n}{X^T} \rightarrow \underset{p \times n}{V^T X^T} \rightarrow \underbrace{\underset{n \times p}{XV}}_{\text{Data in V Basis}}$$



Note $V_1 \perp V_2$

V = rotation

Change basis



$\mathrm{Cov}(\tilde{X}) = \Sigma^2$!

diagonal

No covariances!

"decorrelated"

Check

$$\mathrm{Cov}(\tilde{X}) = \tilde{X}^T X = (XV)^T XV = V^T X^T X V$$
$$= V^T V \Sigma^2 V^T V = \Sigma^2$$
$$\Sigma^2 = \mathrm{diag}(\sigma_1^2, \sigma_2^2, \ldots, \sigma_p^2)$$

$$\underset{n\times p}{X} \longrightarrow \underset{n\times p}{\tilde{X}} = XV$$

correlated

decorrelated

Variances $\sigma_1^2 \geq \sigma_2^2 \geq \dots \geq \sigma_p^2$

Largest to smallest

$$V = [\,V_1 \;\; V_2 \; \dots \; V_p\,]$$

V_i = i^{th} Principal Component

σ_i^2 = Amt of Variance in ith Principal Component

In Practice

Use the first few (2, 3, 4) Principal Components

First k: $V_1, V_2 \dots V_k$ Total Variance explained

$\sigma_1^2 + \sigma_2^2 + \dots + \sigma_k^2 \leftarrow$

Interpreting Principal Components

Loadings

$G_1 \quad G_2 \quad \dots \quad G_p$ Predictors

V: Original Predictors $\longrightarrow$ Principal Comp
$G_1, G_2 \dots G_p \qquad\qquad V_1\ V_2 \dots V_p$

$$V = \begin{array}{c} \\ G_1 \\ G_2 \\ G_3 \\ \vdots \\ G_p \end{array} \begin{array}{c} \begin{array}{cccc} V_1 & V_2 & \dots & V_p \end{array} \\ \left[\begin{array}{cccc} & & & \\ & & & \\ L_{31} & L_{32} & & \\ & & & \\ & & & \end{array}\right] \end{array}$$

Loading of G_3 on P_1, P_2

Use 1st 2
Principal Components

Loadings: If $L_{31} >> L_{32}$
G_3 is heavily "loaded" on V_1

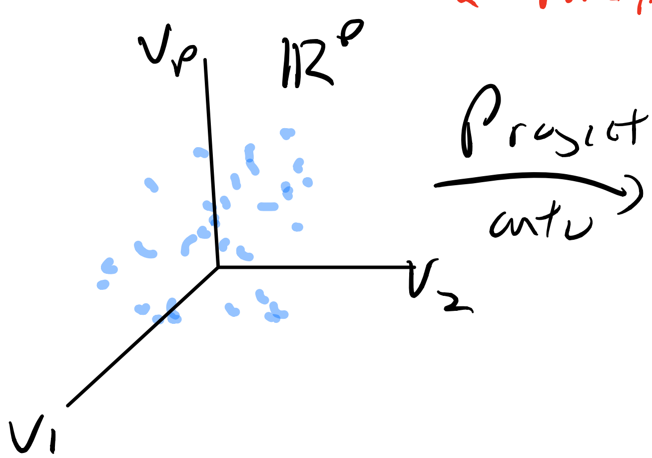
Interpreting Principal Component

Data

$$
V = \begin{array}{c} \\ a_1 \\ a_2 \\ a_3 \\ \vdots \\ a_p \end{array}
\begin{array}{c} \begin{array}{cccc} v_1 & v_2 & \cdots & v_p \end{array} \\ \left[\begin{array}{cccc} \boxed{\begin{array}{cc} & \\ & \\ & \\ & \\ & \end{array}} & & & \end{array} \right] \end{array}
$$

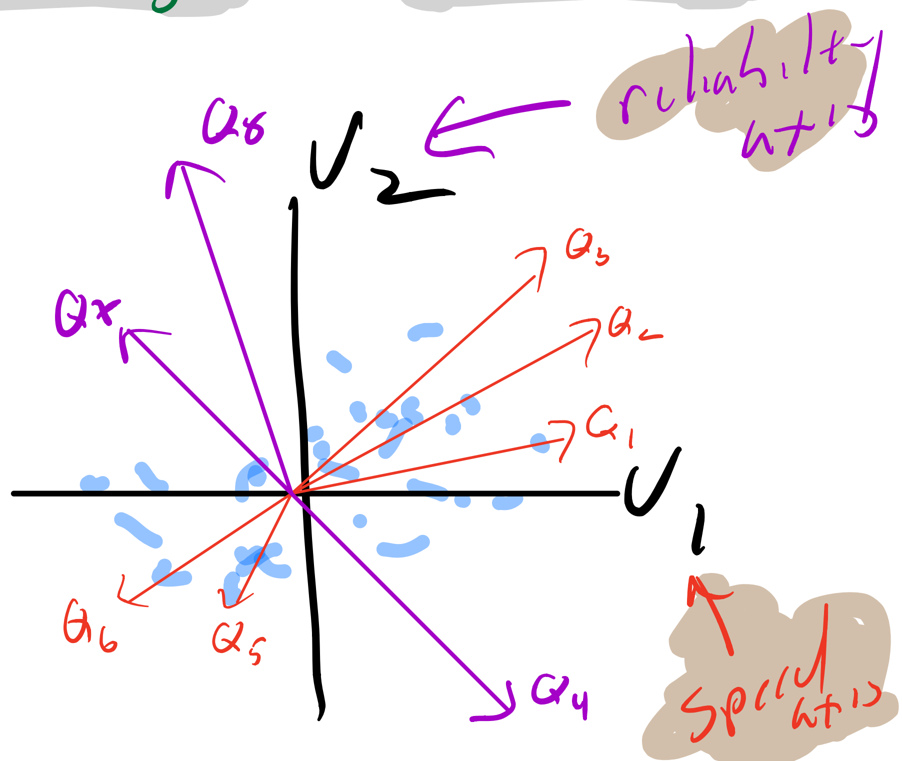
Use 1st 2 Principal Components $= V_{12}$

$\tilde{X}_{12} = X V_{12} \equiv$ Projection of Rotated data in $\mathbb{R}^p$ onto 1st 2 Principal Components



Interpreting Principal Component

Loadings Data



$Q_1 Q_2 Q_3\ Q_5 Q_6 \equiv$ Shared Qualities
"reliability"?

$Q_4 Q_7 Q_8 \equiv$ Different Shared Quality
"speed"?

Observations near $\begin{cases} V_1 \text{ axis: Speed dominates} \\ V_2 \text{ axis: Reliability dominates} \end{cases}$