Principal Component Analysis (PCA)

Consider the eigendecomposition of the covariance matrix

unitary ($\underline{Q}\,\underline{Q}^T = \underline{Q}^T\underline{Q} = \underline{I}$)

$$\underline{V} = \underline{Q}\,\underline{\Lambda}\,\underline{Q}^T$$

diagonal matrix of real, non-negative eigenvalues

covariance matrix (normal, symmetric, positive semi-definite)

$$\underline{V} = \begin{bmatrix} \underline{q}_1 & \underline{q}_2 & \cdots & \underline{q}_m \end{bmatrix} \begin{bmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_m \end{bmatrix} \begin{bmatrix} \underline{q}_1^T \\ \underline{q}_2^T \\ \vdots \\ \underline{q}_m^T \end{bmatrix} \quad \text{with } \|\underline{q}_i\| = 1$$

$$= \begin{bmatrix} \underline{q}_1 & \underline{q}_2 & \cdots & \underline{q}_m \end{bmatrix} \begin{bmatrix} \lambda_1 \underline{q}_1^T \\ \lambda_2 \underline{q}_2^T \\ \vdots \\ \lambda_m \underline{q}_m^T \end{bmatrix} = \lambda_1 \underline{q}_1 \underline{q}_1^T + \lambda_2 \underline{q}_2 \underline{q}_2^T + \ldots + \lambda_m \underline{q}_m \underline{q}_m^T$$

This is similar to SVD, which can be written as a summation of rank-1 matrices

Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m \geq 0$

Then $\underline{q}_1$ contains the direction with the most variation, then $\underline{\varphi}_1$ contains the direction with the most variation, because λ_1 is the largest eigenvalue

The $\underline{\varphi}_i$ vectors are called principal directions

We can obtain principal components (or scores) by multiplying the φ matrix by the original data

$$\text{scores} = \underline{D}\,\underline{\varphi} \qquad \underline{D}: \text{data used to obtain } \underline{V}$$

$\Rightarrow$ These can identify outliers in the observations

Methods associated with principal directions are called Principal Component Analysis (PCA)

Typically data is shifted to have zero mean by defining

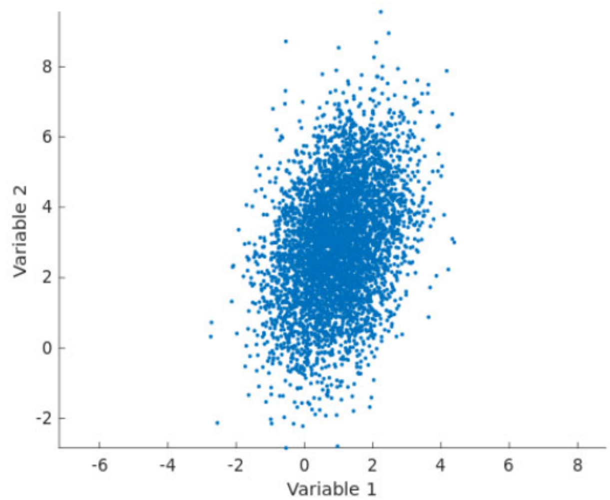
$$\hat{x}_i = x_i - \mu,$$

which lets one compare widely varying data sets

Example: Consider two random variables w/ $\mu = 1$, $\mu = 2$

Example: Consider two random variables w/ $\mu_1 = 1$, $\mu_2 = 3$

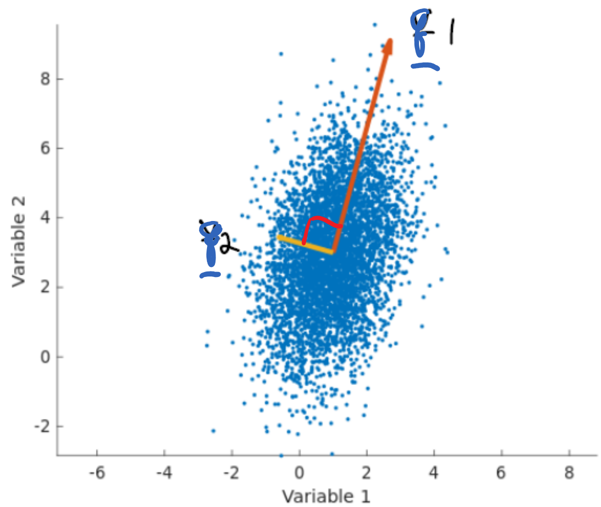
and $\underline{V} = \begin{bmatrix} 1 & 3/5 \\ 3/5 & 3 \end{bmatrix}$, $N = 5000$ (data points)



Eigensystem

$\lambda_1 = 3.1662$ $\quad \underline{q}_1 = \begin{bmatrix} 0.2669 \\ 0.9637 \end{bmatrix}$ $\quad \underline{q}_1^T \underline{q}_1 = 1$

$\lambda_2 = 0.8338$ $\quad \underline{q}_2 = \begin{bmatrix} -0.9637 \\ 0.2669 \end{bmatrix}$ $\quad \underline{q}_1^T \underline{q}_2 = 0$

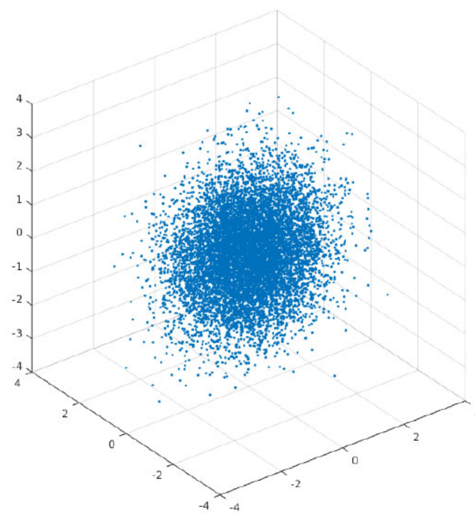


Note: $\underline{q}_1 \perp \underline{q}_2$

Example:

3-dimensional problem ($m = 3$)

with $N = 10{,}000$



$$\bar{\underline{x}} = \begin{bmatrix} -0.0005 \\ -0.0023 \\ -0.0124 \end{bmatrix}$$

$$\underline{V} = \begin{bmatrix} 1.0061 & & \text{sym} \\ 0.1026 & 0.9764 & \\ 0.2102 & 0.0096 & 0.9932 \end{bmatrix}$$

Eigensystem:

$$\begin{bmatrix} 0.7135 \end{bmatrix}$$

Eigensystem:

$$\lambda_1 = 1.2357 \qquad \underline{q}_1 = \begin{bmatrix} 0.7135 \\ 0.3056 \\ 0.6306 \end{bmatrix}$$

$$\lambda_2 = 0.9721 \qquad \underline{q}_2 = \begin{bmatrix} -0.0024 \\ -0.8985 \\ 0.4389 \end{bmatrix}$$

$$\lambda_3 = 0.7679 \qquad \underline{q}_3 = \begin{bmatrix} 0.7006 \\ -0.3150 \\ -0.6402 \end{bmatrix}$$

% of information in each component: $\dfrac{\lambda_i}{\sum \lambda_j}$

λ_1: 41.5%, λ_2: 32.7%, λ_3: 25.8%

Using $\underline{q}_1$ + $\underline{q}_2$ only, 74.2% of all information is in the plane given by $\underline{q}_1$ + $\underline{q}_2$

Use this idea for compression!

Example: Outliers

There are four students taking three exams

Each student is an observation ($N = 4$), while each exam is the data ($m = 3$)

	Math (m)	Physics (p)	Chemistry (c)
S1	90	70	90
S2	80	80	75
S3	70	80	80
S4	75	70	75

PCA

1. Determine the mean of each piece of data

$$\bar{m} = 78.75, \quad \bar{p} = 75, \quad \bar{c} = 80$$

2. Center the data $x_i - \bar{x}$

	$\hat{m}$	$\hat{p}$	$\hat{c}$
S1	11.25	-5	10
S2	1.25	5	-5
S3	-8.75	5	0
S4	-3.75	-5	-5
Σ	0.00	0	0

✓

3. Compute covariance matrix for data

$$\sigma_{mp} = \frac{1}{N-1} \sum_{i=1}^{N} \left(\hat{m}_i - \cancelto{0}{\bar{\hat{m}}}\right)\left(\hat{p}_i - \cancelto{0}{\bar{\hat{p}}}\right) = \frac{1}{N-1} \sum_{i=1}^{N} \hat{m}_i \hat{p}_i$$

$$= -\frac{75}{3} = -25$$

and similarly for other components of $\underline{V}$

$$\underline{V} = \begin{bmatrix} 72.917 & - & \text{sym} \\ -25 & 33.333 & | \\ 41.667 & -16.667 & 50 \end{bmatrix}$$

④ Compute eigenvalues & eigenvectors of $\underline{V}$

$$\lambda_1 = 115.61 \qquad \underline{q}_1 = \begin{bmatrix} 0.7513 \\ -0.3426 \\ 0.5642 \end{bmatrix}$$

$$\lambda_2 = 23.063 \qquad \underline{q}_2 = \begin{bmatrix} -0.0471 \\ -0.8804 \\ -0.4719 \end{bmatrix}$$

$$\lambda_3 = 17.581 \qquad \underline{q}_3 = \begin{bmatrix} 0.6583 \\ 0.3280 \\ -0.6775 \end{bmatrix}$$

Note: $\lambda_1, \underline{q}_1$ contains $\dfrac{115.61}{115.61 + 23.063 + 17.581} \approx 0.74$

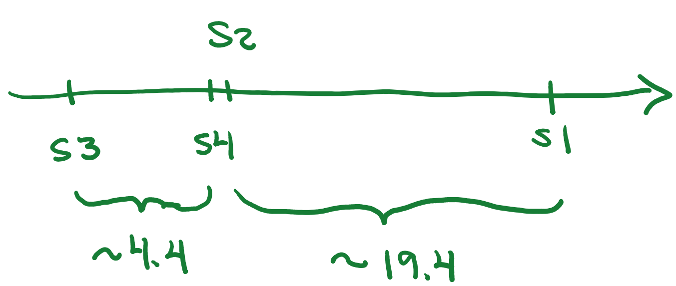
74% of information content

⑤ Calculate the scores $\underline{D}\,\underline{Q}$ ($\underline{D}$ contains shifted data)

$$\begin{bmatrix} 11.25 & -5 & 10 \\ 1.25 & 5 & -5 \\ -8.75 & 5 & 0 \\ -3.75 & -5 & -5 \end{bmatrix} \begin{bmatrix} 0.7513 & -0.0471 & 0.6583 \\ -0.3426 & -0.8804 & 0.3280 \\ 0.5642 & -0.4719 & -0.6775 \end{bmatrix}$$

$$= \begin{array}{cc} \begin{bmatrix} 15.8054 & -0.8467 & -1.0088 \\ -3.5945 & -2.1011 & 5.8504 \\ -8.2863 & -3.9901 & -4.1205 \\ -3.9251 & 6.9380 & -0.7211 \end{bmatrix} & \begin{matrix} S1 \\ S2 \\ S3 \\ S4 \end{matrix} \\ \begin{matrix} \boxed{PC\,1} & \quad PC\,2 & \quad PC\,3 \end{matrix} & \end{array}$$

⑥ Plot PC1



Clearly, student S1 is outlier!

PCA is useful to identify dominant components and to reduce data

---

Aside: Direct connection to SVD is not obvious from above presentation

Let $\underline{A} = \underline{D}^T$, so that $\underline{A} \in \mathbb{R}^{m \times n}$

Then, covariance matrix is simply $\frac{1}{n-1}\underline{A}\,\underline{A}^T$

Let's rename this $\underline{K} = \frac{1}{n-1}\underline{A}\,\underline{A}^T$ (rather than $\underline{V}$)

Clearly $\underline{K}$ is normal, so that $\underline{K} = \underline{Q}\,\underline{\Lambda}\,\underline{Q}^T$ with unitary $\underline{Q}$ and diagonal $\underline{\Lambda}$

Furthermore, find SVD of $\underline{A}$

$$\underline{A} = \underline{U}\,\underline{\Sigma}\,\underline{V}^T \quad \text{with unitary } \underline{U} + \underline{V}$$

and

$$\underline{A}^T = \underline{V}\,\underline{\Sigma}^T\underline{U}^T$$

such that

$$\underline{K} = \frac{1}{n-1}\underline{A}\,\underline{A}^T = \underline{U}\,\underline{\Sigma}\,\underline{V}^T\underline{V}\,\underline{\Sigma}^T\underline{U}^T$$

$$\underline{K} = \underline{U}\,\underline{\Sigma}^2\,\underline{U}^T \quad \text{with} \quad \underline{U} \leftrightarrow \underline{Q}$$
$$\underline{\Sigma}^2 \leftrightarrow \underline{\Lambda}$$

Then, the scores $\underline{D}\,\underline{Q}$ become simply $\underline{A}^T\underline{U}$

$$\underline{A}^T\underline{U} = \underline{V}\,\underline{\Sigma}^T\,\underline{U}^T\,\underline{U}$$

$$\underline{A}^T\underline{U} = \underline{V}\,\underline{\Sigma}$$

$\therefore$ Just perform SVD on shifted data to evaluate scores from $\underline{D}\,\underline{Q} = \underline{V}\,\underline{\Sigma}$

Issue: Need to have covariance matrix, which is not given in most situations

One possible approach:

Monte Carlo Methods

Random sampling → mean calculation → covariance matrix

Example: Flipping of two coins, each with $p = 1/2$

Example: Flipping of two coins, each with $p = 1/2$

Algorithm

Choose N trials

for $i = 1:N$

% coin #1

r = random number in $[0, 1]$

$$x_i = \begin{cases} 0 & \text{if } r < 0.5 \quad \text{\% tails} \\ 1 & \text{if } r \geq 0.5 \quad \text{\% heads} \end{cases}$$

% coin #2

r = random number in $[0, 1]$

$$y_i = \begin{cases} 0 & \text{if } r < 0.5 \quad \text{\% tails} \\ 1 & \text{if } r \geq 0.5 \quad \text{\% heads} \end{cases}$$

end

$$\left.\begin{aligned} \bar{x} &= \text{sum}(x)/N \\ \bar{y} &= \text{sum}(y)/N \end{aligned}\right\} \text{mean}$$

$V = \text{zeros}(2, 2)$ % covariance matrix for $i = 1:N$

$$V_{11} = V_{11} + (x_i - \bar{x})^2$$

$$V_{12} = V_{12} + (x_i - \bar{x})(y_i - \bar{y})$$

$$V_{22} = V_{22} + (y_i - \bar{y})^2$$

end

$$V_{11} = V_{11} / (N-1)$$

$$V_{12} = V_{12} / (N-1)$$

$$V_{21} = V_{12}$$

$$V_{22} = V_{22} / (N-1)$$

In general, there are multiple inputs + outputs

for $i = 1:N$

select $x_0, x_1, \ldots, x_m$ from appropriate distribution

e.g., x_0 from uniform, x_1 from Gaussian, ...

use $x_0 \rightarrow x_m$ in the model, which results in outputs $y_0 \to y_n$

Issue: This is a brute force method, which is very slow to converge

Modern sophisticated techniques are much faster:

- Quasi Monte Carlo (modify distributions to cover probability space faster)
- Importance Sampling
- Latin Hypercube
- Others (see Dakota, Queso)