

Goals

- principal component analysis
- data = point cloud
- variance & covariance
- orthogonal diagonalization

Suppose each trial of an experiment makes m real-valued measurements, and we run n trials. This produces

$$P = \underbrace{\{x_1, \dots, x_n\}}_{\text{point cloud}} \in \mathbb{R}^m$$

point cloud

The mean of P is $\mu := \frac{1}{n} (x_1 + \dots + x_n)$.

The recentering of P is $B = \begin{pmatrix} x_1 - \mu & x_2 - \mu & \dots & x_n - \mu \end{pmatrix} \in \mathbb{R}^{m \times n}$.

Note The mean of the columns of B is $\vec{0}$.

Defn The covariance matrix of P is the symmetric matrix.

$$S = \frac{1}{n-1} BB^T \in \mathbb{R}^{m \times m}$$

$$(BB^T)^T = (B^T)^T \cdot B^T = BB^T$$

Note $S_{ii} = \frac{1}{n-1} \sum_{j=1}^n (x_{ji} - \mu_i)^2 = \text{variance of } i\text{-th measurement/variable}$

For $i \neq j$, $S_{ij} = \frac{1}{n-1} \sum_{j=1}^n (x_{ji} - \mu_i)(x_{ij} - \mu_j) = \text{covariance of } i\text{-th, } j\text{-th variables}$

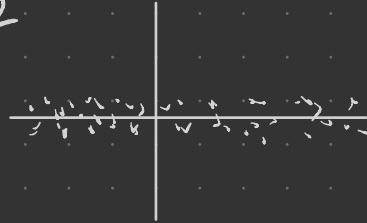
- Variance measures how much things differ from the mean.

Its square root is standard deviation.

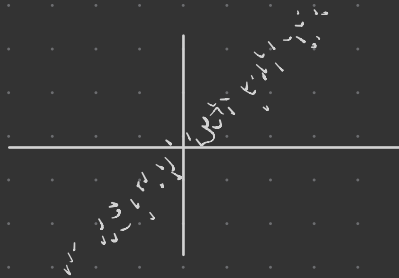
- Covariance measures how variables depend on each other.

Defn The total variance of P is $\text{tr}(S) = \sum \text{variance of variables}$.

E.g. $m=2$



$$S = \begin{pmatrix} 95 & 1 \\ 1 & 5 \end{pmatrix}$$



$$S = \begin{pmatrix} 50 & 40 \\ 40 & 50 \end{pmatrix}$$

Idea Use linear algebra to recognize the similarity of these data.

Spectral Thm If $A \in \mathbb{R}^{n \times n}$ and $A = A^T$, then $\exists \lambda_1, \dots, \lambda_n \in \mathbb{R}$ and orthonormal nonzero vectors $v_1, \dots, v_n$ such that $Av_i = \lambda v_i$.

Note · If $P = \begin{pmatrix} | & & | \\ v_1 & \dots & v_n \\ | & & | \end{pmatrix}$ then $PP^T = I_n$ and

$$\text{diag}(\lambda_1, \dots, \lambda_n) = P^{-1}AP.$$

- For $B \in \mathbb{R}^{m \times n}$, $BB^T = (BB^T)^T \in \mathbb{R}^{m \times m}$ and the spectral theorem applies.

Prop BB^T and B^TB share the same nonzero eigenvalues.
 $\overset{\mathbb{R}^{m \times m}}{\uparrow}$ $\overset{\mathbb{R}^{n \times n}}{\uparrow}$

PF Take v an eigenvector of B^TB with eigenvalue $\lambda \neq 0$, so

$$B^TBv = \lambda v$$

$$\Rightarrow (BB^T)(Bv) = \lambda(Bv) \quad [\text{mult on left by } B]$$

$\Rightarrow \lambda$ is an eigenvalue of BB^T with eigenvector Bv .
 $(Bv \neq 0 \text{ b/c } B^TBv = \lambda v \neq 0)$

Similarly, $BB^Tw = \lambda w \Rightarrow B^TB(B^Tw) = \lambda(B^Tw)$, so BB^T, B^TB have the same nonzero eigenvalues. $\square$

Q How should we find eigenvalues when $B \in \mathbb{R}^{500 \times 2}$?

A $BB^T \in \mathbb{R}^{500 \times 500}$

$$B^TB \in \mathbb{R}^{2 \times 2}$$

work on this $\chi_{B^TB}(x)$ quadratic polynomial

$\leadsto \lambda_1, \lambda_2$ eigenvals of B^TB

BB^T has eigenvalues $\lambda_1, \lambda_2, 0$.

Prop The eigenvalues of BB^T , B^TB are all nonnegative.

Pf Take v an eigenvector of B^TB w/ eigenvalue λ . Then

$$\begin{aligned}\|Bv\|^2 &= Bv \cdot Bv \\ &= (Bv)^T (Bv) \\ &= v^T (B^T B) v \\ &= v^T (\lambda v) \\ &= \lambda (v^T v) \\ &= \lambda \|v\|^2.\end{aligned}$$

Since $\|Bv\|^2 \geq 0$ and $\|v\|^2 > 0$,
must have $\lambda \geq 0$. $\square$

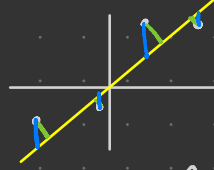
The covariance matrix $S = \frac{1}{n-1} BB^T$ thus has nonnegative real eigenvalues $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_m \geq 0$.

Let $u_1, \dots, u_m$ be the corresponding orthonormal eigenvectors, these are the principal components of the data.

Note • total variance = $\text{tr}(S) = \lambda_1 + \dots + \lambda_m =: T$

- The first principal component u_1 accounts for $\frac{\lambda_1}{T}$ of the total variance. In general, u_i accounts for $\frac{\lambda_i}{T}$ of the total variance.
- u_1 points in the "most significant" direction.
- Among $u_1^\perp$, u_2 points in the most significant direction. Etc.

Fact span $\{u_1\}$ minimizes orthogonal distance from line to point cloud.



least squares
PCA

E.g. 150 iris flowers measured by sepal and petal length/width

