Recap

- (Ch 1-2) Looking at data and relationships
- Linear algebra review

Today

- (Ch 10) Data in high dimensions
  - Visualizing data
  - Summarizing data
  - Dimensionality reduction

Eigenvalues and eigenvectors review

- If A is an $n \times n$ square matrix, an eigenvalue λ and its corresponding eigenvector $\boldsymbol{v}$ (of dimension $n \times 1$) have the property that $A\boldsymbol{v} = \lambda\boldsymbol{v}$
- To solve for λ, we solve the characteristic equation $|A - \lambda I| = 0$
- Given a value of λ, we find the corresponding eigenvector(s) by solving $(A - \lambda I)\boldsymbol{v} = 0$
- Note that if $\boldsymbol{v}$ is an eigenvector for λ, then so is any multiple $k\boldsymbol{v}$

Eigenvalues and eigenvectors: example

Find the eigenvalues and eigenvectors of $A = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$

$$|A - \lambda I| = \begin{vmatrix} 5-\lambda & 3 \\ 3 & 5-\lambda \end{vmatrix} = (5-\lambda)^2 - 3^2 = \lambda^2 - 10\lambda + 25 - 9$$

$$= \lambda^2 - 10\lambda + 16 = (\lambda - 8)(\lambda - 2) = 0$$

So eigenvalues are $\lambda_1 = 8$ and $\lambda_2 = 2$

For $\lambda_1 = 8$

$$A - 8I = \begin{bmatrix} 5-8 & 3 \\ 3 & 5-8 \end{bmatrix} = \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 \\ 1 & -1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & -1 \\ 0 & 0 \end{bmatrix}$$

So $v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$

For $\lambda_2 = 2$

$$A - 2I = \begin{bmatrix} 5-2 & 3 \\ 3 & 5-2 \end{bmatrix} = \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix}$$

So $v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$

Diagonalization of a symmetric matrix

- If A is an $n \times n$ **symmetric** square matrix, the eigenvalues are real
- If the eigenvalues are also distinct, their eigenvectors are orthogonal
- We can then scale the eigenvectors $\boldsymbol{v}_i$ to ones of unit length $\boldsymbol{u}_i$ and place them into an orthogonal matrix $U = [\boldsymbol{u}_1\ \boldsymbol{u}_2\ \ldots\ \boldsymbol{u}_n]$
- Then we can write a diagonal matrix $\Lambda = U^T A U$ such that the diagonal entries of Λ are $\lambda_1, \lambda_2, \ldots, \lambda_n$ in that order

Diagonalization example

Diagonalize $A = \begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}$

$$\lambda_1 = 8 \Rightarrow v_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \Rightarrow u_1 = \frac{1}{\|v_1\|} v_1 = \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\lambda_2 = 2 \Rightarrow v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix} \Rightarrow u_2 = \frac{1}{\|v_2\|} v_2 = \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\underbrace{\begin{bmatrix} 8 & 0 \\ 0 & 2 \end{bmatrix}}_{\Lambda} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{U^T} \underbrace{\begin{bmatrix} 5 & 3 \\ 3 & 5 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}}_{U}$$

Visualizing multidimensional data

- Visualizations (see today's Jupyter notebook)
  - 3D scatter plot (for 3-dimensional data)
  - Scatter plot matrix for 3 or more dimensions

- Reducing the dimensionality before visualization

Summarizing multidimensional data

- We need location and spread parameters for multidimensional data

- Notation
  - Suppose the dataset $\{\mathbf{x}\}$ consists of N items
  - Each item $\mathbf{x}_i$ is a d-dimensional vector of numbers
  - We refer to the jth component of the ith item as $\mathbf{x}_i^{(j)}$

Mean of a multidimensional dataset

- We compute the mean of $\{\mathbf{x}\}$ by computing the means of each component separately and stacking them into a vector

$$\text{mean of } j\text{th component} = \frac{\sum_i \mathbf{x}_i^{(j)}}{N}$$

- We write the mean of $\{\mathbf{x}\}$ as

$$\text{mean}(\{\mathbf{x}\}) = \frac{\sum_i \mathbf{x}_i}{N}$$

Covariance of a pair of components

- Recall that the covariance of random variables X and Y is

$$\mathrm{cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$$

- For a dataset with a pair of components $\{\mathbf{x}^{(j)}\}$ and $\{\mathbf{x}^{(k)}\}$

$$\mathrm{cov}\left(\{\mathbf{x}^{(j)}\}, \{\mathbf{x}^{(k)}\}\right) = \frac{\sum_i (\mathbf{x}_i^{(j)} - \mathrm{mean}(\{\mathbf{x}^{(j)}\}))(\mathbf{x}_i^{(k)} - \mathrm{mean}(\{\mathbf{x}^{(k)}\}))}{N}$$

Properties of covariance

- The covariance of a component with itself is its variance

$$\operatorname{cov}\left(\{\mathbf{x}^{(j)}\},\{\mathbf{x}^{(j)}\}\right) = \operatorname{var}\left(\{\mathbf{x}^{(j)}\}\right) = \operatorname{std}\left(\{\mathbf{x}^{(j)}\}\right)^2$$

- The correlation coefficient is the covariance scaled by standard deviations of each component

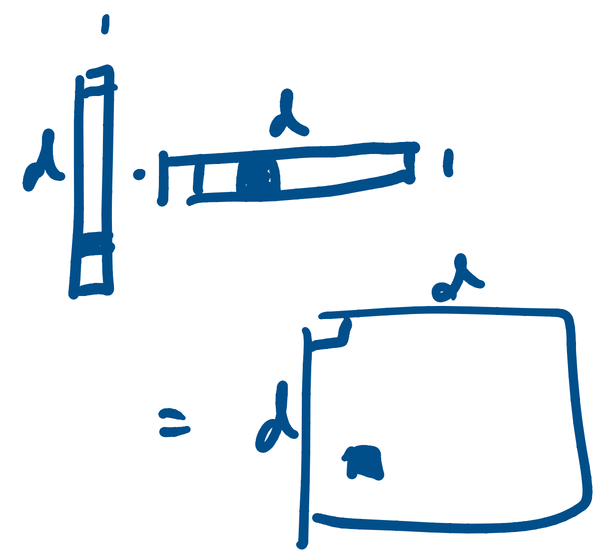
$$\operatorname{corr}\left(\{\mathbf{x}^{(j)},\mathbf{x}^{(k)}\}\right) = \frac{\sum_i \widehat{\mathbf{x}_i^{(j)}}\widehat{\mathbf{x}_i^{(k)}}}{N} = \frac{\operatorname{cov}\left(\{\mathbf{x}^{(j)}\},\{\mathbf{x}^{(k)}\}\right)}{\operatorname{std}(\{\mathbf{x}^{(j)}\})\operatorname{std}(\{\mathbf{x}^{(k)}\})}$$

Covariance matrix of multidimensional data

- We can capture all the pairwise covariances in a $d \times d$ matrix

$$\text{Covmat}(\{\mathbf{x}\}) = \frac{\sum_i (\mathbf{x}_i - \text{mean}(\{\mathbf{x}\}))(\mathbf{x}_i - \text{mean}(\{\mathbf{x}\}))^T}{N}$$

- Properties
  - The (j, k) entry of $\text{Covmat}(\{\mathbf{x}\})$ is $\text{cov}\left(\{\mathbf{x}^{(j)}\}, \{\mathbf{x}^{(k)}\}\right)$
  - The (j, j) entry of $\text{Covmat}(\{\mathbf{x}\})$ is $\text{var}\left(\{\mathbf{x}^{(j)}\}\right)$
  - $\text{Covmat}(\{\mathbf{x}\})$ is symmetric



Translation properties

- Translating the data translates the mean

$$\text{mean}(\{\mathbf{x} + \mathbf{c}\}) = \text{mean}(\{\mathbf{x}\}) + \mathbf{c}$$

- Translating the data leaves the covariance matrix unchanged

$$\text{Covmat}(\{\mathbf{x} + \mathbf{c}\}) = \text{Covmat}(\{\mathbf{x}\})$$

Proof

$$\text{Covmat}(\{\mathbf{x}+\mathbf{c}\})$$

$$= \frac{\sum_i (\mathbf{x}_i+\mathbf{c}-\text{mean}(\{\mathbf{x}+\mathbf{c}\}))(\mathbf{x}_i+\mathbf{c}-\text{mean}(\{\mathbf{x}+\mathbf{c}\}))^T}{N}$$

$$= \frac{\sum_i (\mathbf{x}_i+\mathbf{c}-\text{mean}(\{\mathbf{x}\})-\mathbf{c})(\mathbf{x}_i+\mathbf{c}-\text{mean}(\{\mathbf{x}\})-\mathbf{c})^T}{N}$$

by mean property

$$= \frac{\sum_i (\mathbf{x}_i-\text{mean}(\{\mathbf{x}\}))(\mathbf{x}_i-\text{mean}(\{\mathbf{x}\}))^T}{N}$$

$$= \text{Covmat}(\{\mathbf{x}\})$$

Linear transformation properties

- Linearly transforming the data linearly transforms the mean

$$\text{mean}(\{A\mathbf{x}\}) = A\,\text{mean}(\{\mathbf{x}\})$$

- Linearly transforming the data changes the covariance matrix

$$\text{Covmat}(\{A\mathbf{x}\}) = A\,\text{Covmat}(\{\mathbf{x}\})\,A^{\boldsymbol{T}}$$

$$\text{var}(\{kx\}) = k^2\,\text{var}(\{x\})$$

Proof

$$\mathrm{Covmat}(\{A\mathbf{x}\})$$

$$= \frac{\sum_i (A\mathbf{x}_i - \mathrm{mean}(\{A\mathbf{x}\}))(A\mathbf{x}_i - \mathrm{mean}(\{A\mathbf{x}\}))^T}{N}$$

$$= \frac{\sum_i (A\mathbf{x}_i - A\ \mathrm{mean}(\{\mathbf{x}\}))(A\mathbf{x}_i - A\ \mathrm{mean}(\{\mathbf{x}\}))^T}{N}$$

by mean property

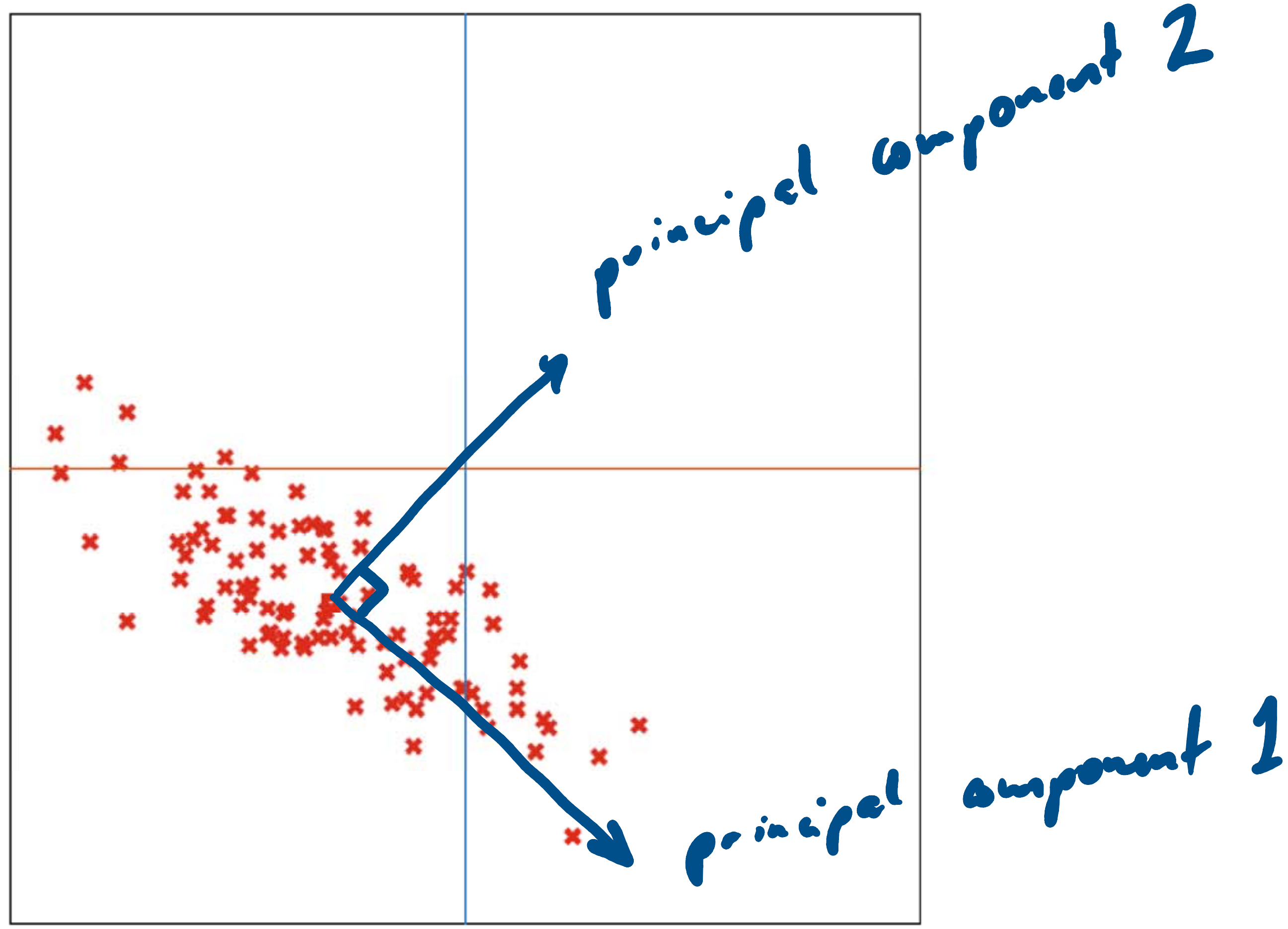
$$= \frac{\sum_i A(\mathbf{x}_i - \mathrm{mean}(\{\mathbf{x}\}))(\mathbf{x}_i - \mathrm{mean}(\{\mathbf{x}\}))^T A^T}{N}$$

$(AB)^T = B^T A^T$

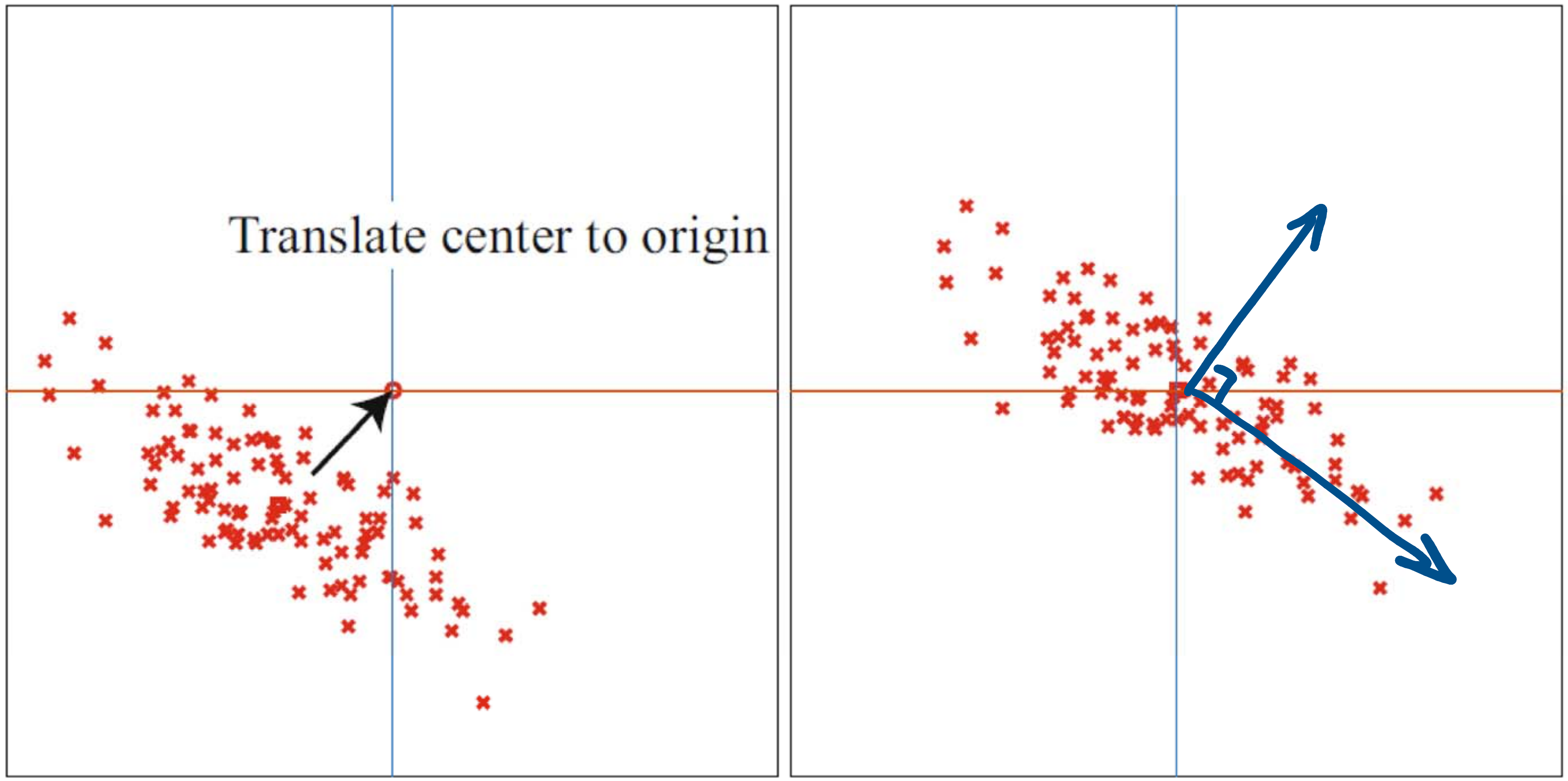
$$= A\,\frac{\sum_i (\mathbf{x}_i - \mathrm{mean}(\{\mathbf{x}\}))(\mathbf{x}_i - \mathrm{mean}(\{\mathbf{x}\}))^T}{N}\,A^T$$

$$= A\ \mathrm{Covmat}(\{\mathbf{x}\})\ A^T$$

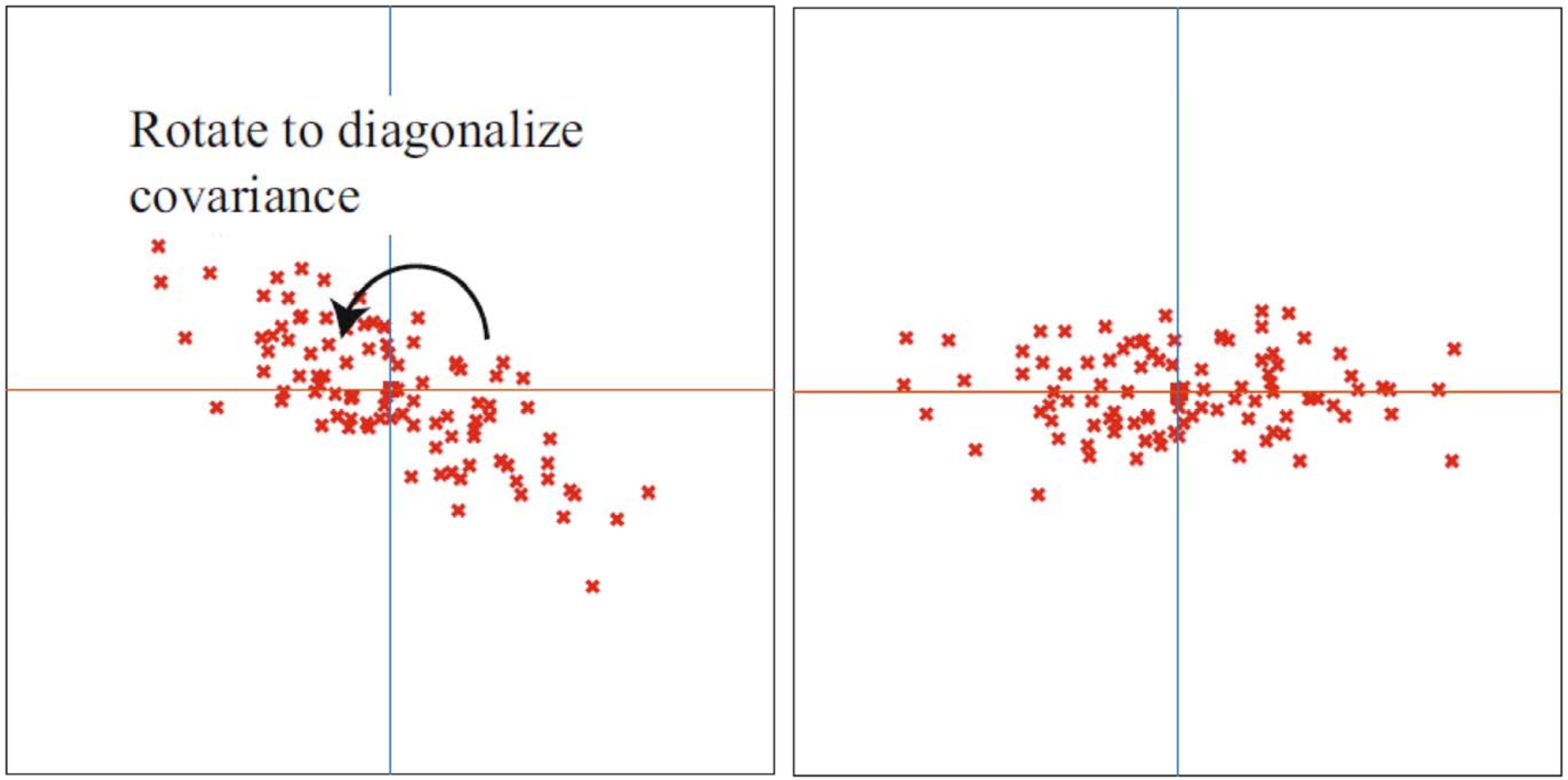
Dimensionality reduction: 2D to 1D example



Step 1: subtract mean



Step 2: apply linear transformation



Step 3: drop component(s)