CS 260 – Foundations of Data Science

Lecture 14 – Principal Component Analysis

10/27/2026

Adam Poliak

Announcements

HW5 due tonight

HW6 due next Tuesday (will be released tonight)

Midterm 2 – November 19th

- propose to change it to Tuesday November 24th

Outline

Linear Algebra: eigenvectors & eigenvalues

Principal Components Analysis (PCA)

Eigenvectors & Eigenvalues

$$\boldsymbol{A}\,\vec{v} = \lambda\,\vec{v}$$

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

Eigenvectors & Eigenvalues

$$\boldsymbol{A}\,\vec{v} = \lambda\,\vec{v}$$

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 3 \end{bmatrix}$$

Determinant

$$A = \begin{bmatrix} 2 & 0 \\ 0 & 3 \end{bmatrix}$$

$$\det(A) = 2(3) - 0(0) = 6$$

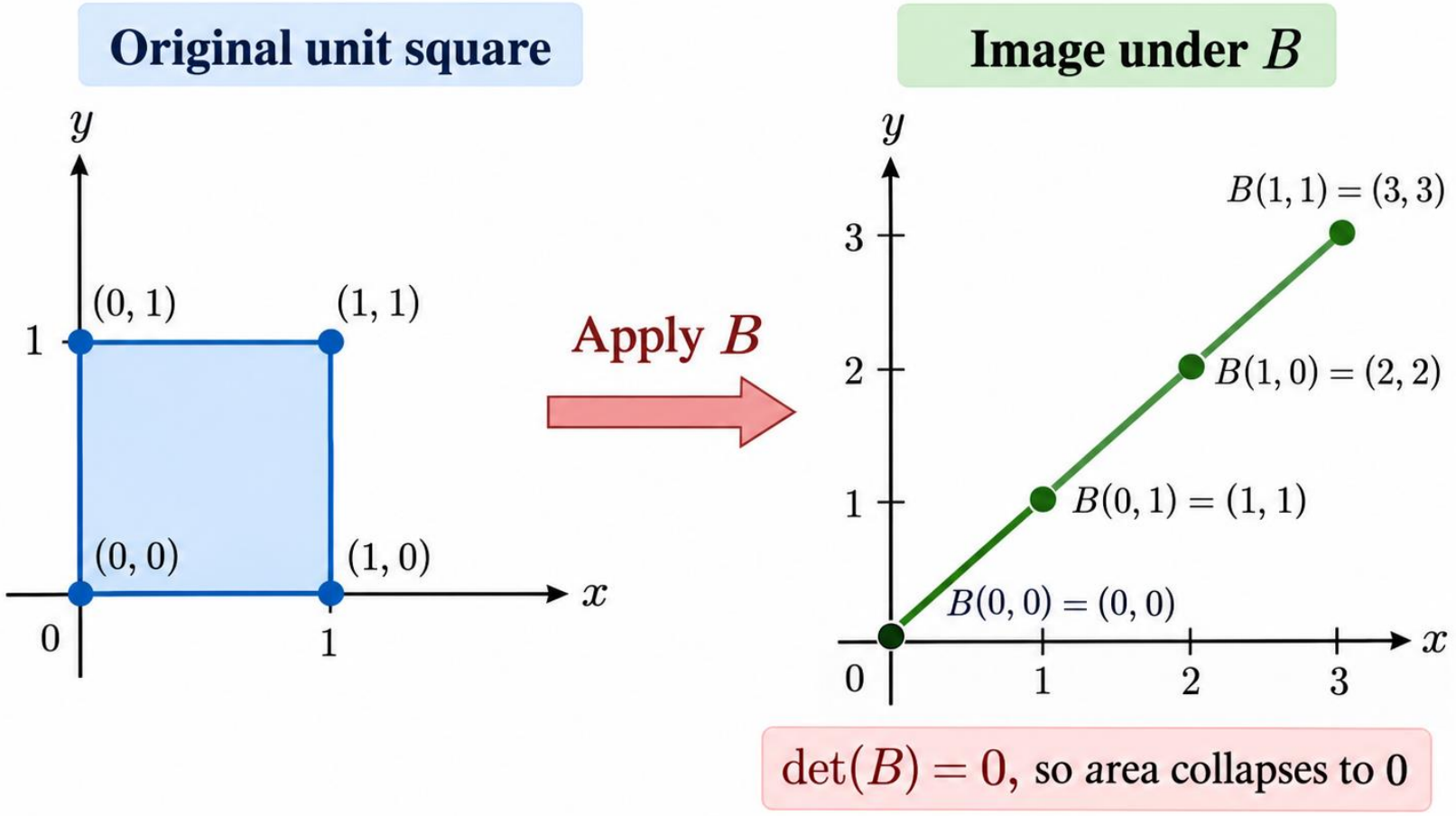
Determinant

$$B = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$$

Determinant

Transforming the unit square by B

$$B = \begin{pmatrix} 2 & 1 \\ 2 & 1 \end{pmatrix}$$



$\det(B) = 0$, so area collapses to 0

Eigenvectors & Eigenvalue

$$\boldsymbol{A}\,\vec{v} = \lambda\,\vec{v}$$

$$\boldsymbol{A}\,\vec{v} - \lambda\,\vec{v} = 0$$

$$(\boldsymbol{A} - \lambda)\,\vec{v} = 0$$

$$\det(A\ - \lambda) = 0$$

Eigenvectors & Eigenvalue

$$\boldsymbol{A}\,\vec{v} = \lambda\,\vec{v}$$

$$\boldsymbol{A}\,\vec{v} - \lambda\,\vec{v} = 0$$

$$(\boldsymbol{A} - \lambda)\,\vec{v} = 0$$

$$\det(A\ - \lambda) = 0$$

Eigenvectors & Eigenvalue

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$$

$$\det(A - \lambda) = 0$$

What happens when our matrix is our data $\mathbf{X}$?

Do we need all the features

Imagine we have 5 features:

Height in centimeters, Height in inches, Arm span, Shoe size, Weight

Are all five features providing independent information?

Could fewer measurements summarize most of the differences between observations?

Why might this be useful for machine learning?

Example

$$X = \begin{bmatrix} 2 & 2 \\ 1 & 1 \\ -1 & -1 \\ -2 & -2 \end{bmatrix}$$

Do we actually need two numbers to locate every observation?

Finding better coordinate system

Why perpendicular?

Why unit vector?

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1,1), \qquad \mathbf{v}_2 = \frac{1}{\sqrt{2}}(1,-1)$$

Project the data

$$X = \begin{bmatrix} 2 & 2 \\ 1 & 1 \\ -1 & -1 \\ -2 & -2 \end{bmatrix}$$

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1,1), \qquad \mathbf{v}_2 = \frac{1}{\sqrt{2}}(1,-1)$$

What happens to the second coordinate for every observation?

Discarding a feature

$$X = \begin{bmatrix} 2 & 2 \\ 1 & 1 \\ -1 & -1 \\ -2 & -2 \end{bmatrix}$$

$$Z = \begin{bmatrix} 2\sqrt{2} & 0 \\ \sqrt{2} & 0 \\ -\sqrt{2} & 0 \\ -2\sqrt{2} & 0 \end{bmatrix}$$

We reduced our data from two dimensions to one without losing any information.

Can we reconstruct the original data

$$\hat{\mathbf{x}} = z_1 \mathbf{v}_1$$

$$Z = \begin{bmatrix} 2\sqrt{2} & 0 \\ \sqrt{2} & 0 \\ -\sqrt{2} & 0 \\ -2\sqrt{2} & 0 \end{bmatrix}$$

$$X = \begin{bmatrix} 2 & 2 \\ 1 & 1 \\ -1 & -1 \\ -2 & -2 \end{bmatrix}$$

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}}(1,1),$$

Would this still work exactly if the points did not all lie on a straight line?

Which direction loses less information?

$$z_1 = (2\sqrt{2}, \sqrt{2}, -\sqrt{2}, -2\sqrt{2})$$
$$z_2 = (0, 0, 0, 0)$$

Projecting along (1,1) keeps the observations distinguishable.
Projecting along (1,-1) collapses every observation to zero.

Which projection would allow us to reconstruct the original observations more accurately?

Measuring preserved information

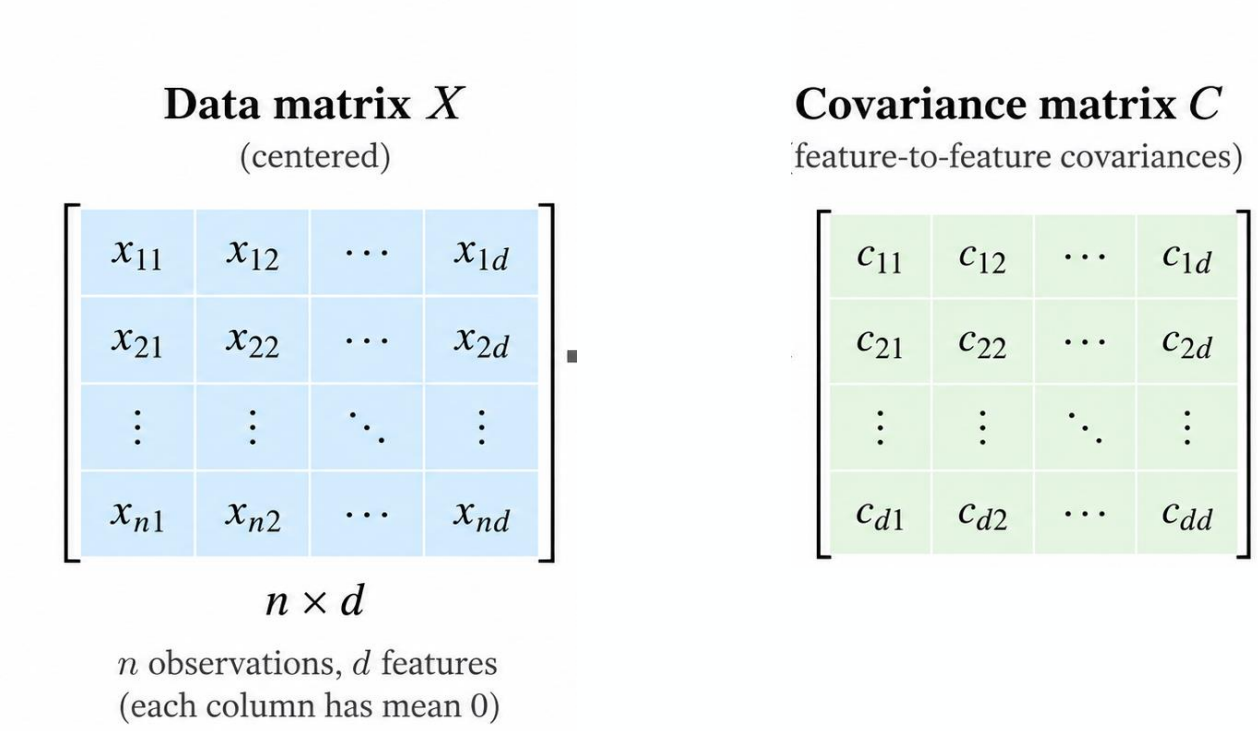
Variance

Goal: find the direction that maximizes variance in the projected data

$$z_1 = (2\sqrt{2}, \sqrt{2}, -\sqrt{2}, -2\sqrt{2})$$
$$z_2 = (0, 0, 0, 0)$$

Covariance

How much does two variables move together linearly



Covariance Matrix

How much does two variables move together linearly

Data matrix X
(centered)

$$\begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1d} \\ x_{21} & x_{22} & \cdots & x_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{nd} \end{bmatrix}$$

$n \times d$

n observations, d features
(each column has mean 0)

Covariance matrix C
feature-to-feature covariances)

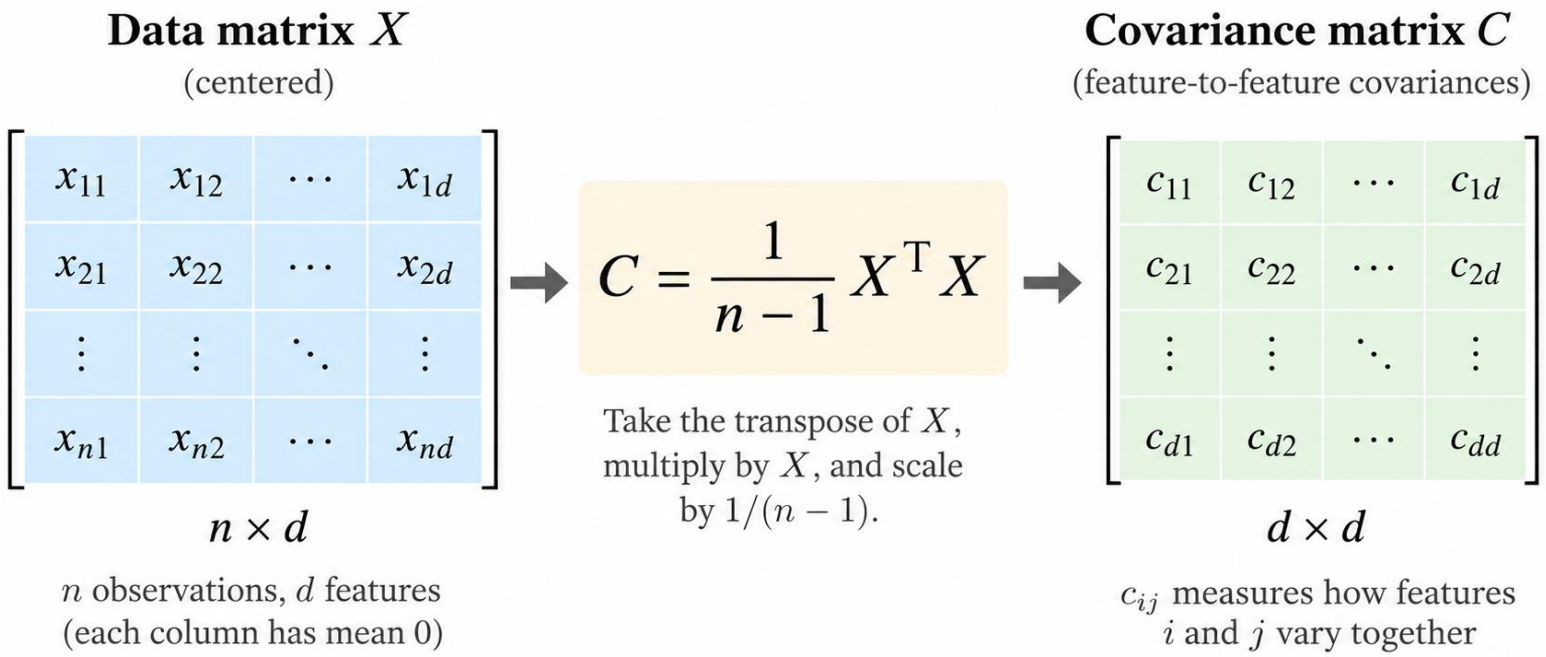
$$\begin{bmatrix} c_{11} & c_{12} & \cdots & c_{1d} \\ c_{21} & c_{22} & \cdots & c_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ c_{d1} & c_{d2} & \cdots & c_{dd} \end{bmatrix}$$

$d \times d$

c_{ij} measures how features
i and j vary together

Covariance Matrix

How much does two variables move together linearly



Covariance and projection

Projected observations: $\mathbf{z} = X\mathbf{v}$

Variance of the project observations: $\mathrm{Var}(\mathbf{z}) = \frac{1}{n-1}\mathbf{z}^T\mathbf{z}$

Eigenvectors & Covariance Matrix

Projected observations: $\mathbf{z} = X\mathbf{v}$

Variance of the project observations: $\mathrm{Var}(\mathbf{z}) = \frac{1}{n-1}\mathbf{z}^T\mathbf{z}$