

Eigenvalues, Eigenvectors, and Principal Component Analysis

Work with a partner. Show your reasoning and calculations. Part I: determinants and computing eigenpairs. Part II: using eigenpairs for PCA. **Part I: Determinants and calculating eigenvectors (Lecture 1)**

(1) From a square to a parallelogram.

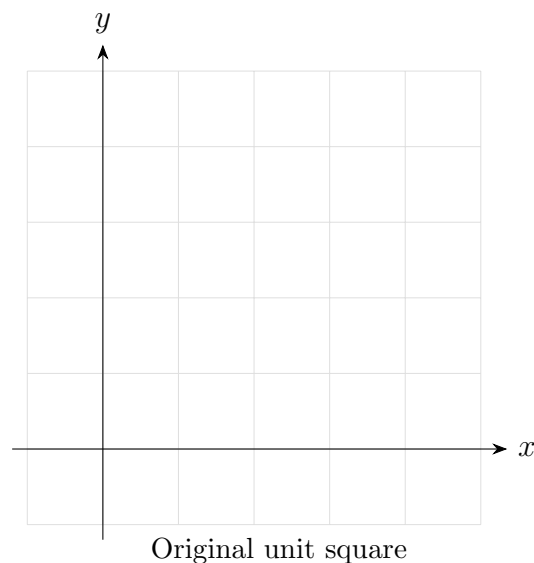
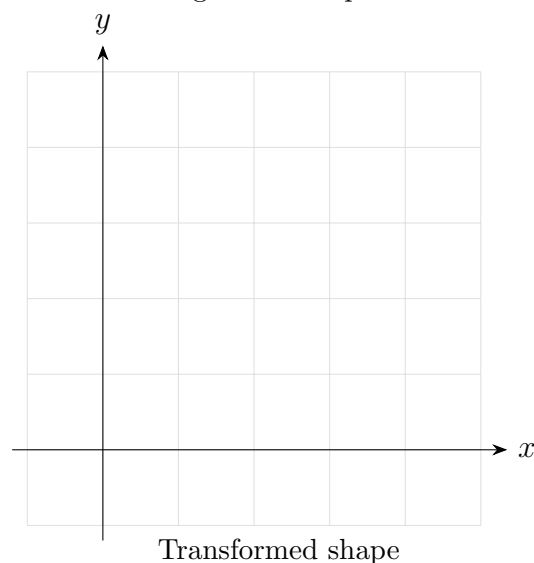
Consider the transformation

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

The unit square has vertices $(0, 0)$, $(1, 0)$, $(1, 1)$, and $(0, 1)$.

(a) Apply A to each of the four vertices. List the transformed coordinates.

(b) Draw the original square and transformed parallelogram on the grids. Label their vertices.



(2) Zero determinants, linear dependence, and lost dimensions.

Let $B = \begin{bmatrix} 2 & 1 \\ 2 & 1 \end{bmatrix}$ and $D = \begin{bmatrix} 2 & 0 \\ 0 & -3 \end{bmatrix}$.

- (a) Which transformation, B or D , collapses area? Which reverses orientation?

- (b) Calculate $\det(B)$ and $\det(D)$.

- (c) Write the two columns of B . Can one column be obtained by scaling the other? Explain why this means the transformed unit square has zero area.

- (d) Find a *nonzero* vector $\mathbf{w}$ such that $B\mathbf{w} = \mathbf{0}$. Rewrite this matrix multiplication as a combination of the columns of B . Why do the terms cancel?

(3) Calculate eigenvalues and eigenvectors.

For this problem, use $M = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix}$.

(a) Write $M - \lambda I$ and expand $\det(M - \lambda I) = 0$. Solve the resulting quadratic equation for both eigenvalues.

(b) For the larger eigenvalue, solve $(M - \lambda I)\mathbf{v} = \mathbf{0}$ for a nonzero eigenvector.

(c) Repeat for the smaller eigenvalue.

(d) Normalize both eigenvectors to have length 1. Check that they are perpendicular.

(4) Your turn: a new matrix.

Let $N = \begin{bmatrix} 3 & 1 \\ 1 & 3 \end{bmatrix}$.

(a) Find both eigenvalues using $\det(N - \lambda I) = 0$.

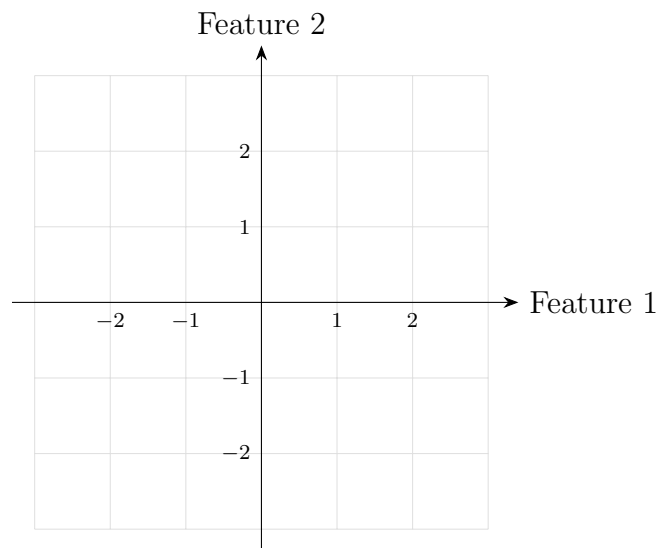
(b) Find a nonzero eigenvector for each eigenvalue, and verify one by calculating $N\mathbf{v}$.

Part II: Using eigenvectors to understand PCA (Lecture 2) Our dataset. Four observations, with two features each, are stored as rows of the matrix

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

(5) Can we describe two features using one number?

(a) Plot the four observations on the axes below. Draw a line through the origin along which they vary.



(b) Why might one number be sufficient to describe each observation? What would that number represent?

(c) Write a nonzero vector $\mathbf{v}_1$ pointing along your line and a vector $\mathbf{v}_2$ perpendicular to it. Check that their dot product is zero.

(d) Convert your two vectors to *unit vectors* (length 1). Why is it helpful to use unit vectors when comparing projections?

(6) Project the observations.

For a unit direction $\mathbf{v}$, the projection coordinate of an observation $\mathbf{x}$ is the dot product $z = \mathbf{x} \cdot \mathbf{v}$. Use the unit vectors

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

(a) Complete the table. Show the dot-product calculation for at least one row.

Observation $\mathbf{x}$	$z_1 = \mathbf{x} \cdot \mathbf{v}_1$	$z_2 = \mathbf{x} \cdot \mathbf{v}_2$
(2, 2)	$2\sqrt{2}$	0
(1, 1)		
(-1, -1)		
(-2, -2)		

(b) If we keep only z_1 and discard z_2 , do we lose any information for this dataset? How does this relate to preserving variation?

(c) Using only z_1 and $\mathbf{v}_1$, reconstruct the original observation (1, 1) with $\hat{\mathbf{x}} = z_1 \mathbf{v}_1$.

(7) What if the data is not perfectly aligned?

$$X_{\text{new}} = \begin{bmatrix} 2 & 2 \\ 1 & 1.5 \\ -1 & -1.5 \\ -2 & -2 \end{bmatrix}.$$

For a unit direction $\mathbf{v}$, the projection coordinate of an observation $\mathbf{x}_{\text{new}}$ is $z_{\text{new}} = \mathbf{x}_{\text{new}} \cdot \mathbf{v}$. Use the unit vectors

$$\mathbf{v}_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad \mathbf{v}_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}.$$

(a) Complete the table. Show the dot-product calculation for at least one row.

Observation $\mathbf{x}_{\text{new}}$	$z_{\text{new},1} = \mathbf{x}_{\text{new}} \cdot \mathbf{v}_1$	$z_{\text{new},2} = \mathbf{x}_{\text{new}} \cdot \mathbf{v}_2$
(2, 2)	$2\sqrt{2}$	0
(1, 1.5)		
(-1, -1.5)		
(-2, -2)		

(b) If we keep only $z_{\text{new},1}$ and discard $z_{\text{new},2}$, do we lose any information for this dataset? How does this relate to preserving variation?

(c) Using only $z_{\text{new},1}$ and $\mathbf{v}_1$, reconstruct the observation $(1, 1.5)$ using $\hat{\mathbf{x}} = z_{\text{new},1}\mathbf{v}_1$. Compare with the original point.

(8) Which direction preserves more variation?

For centered projected coordinates $z_1, \dots, z_n$, the *sample variance* is

$$\text{Var}(z) = \frac{1}{n-1} \sum_{i=1}^n z_i^2.$$

(a) Calculate the sample variance of the four z_1 coordinates from Problem 6.

(b) Calculate the sample variance of the four z_2 coordinates.

(c) Which direction would you keep if you were allowed only one new feature? Why?

(d) Repeat (a)–(c) for the slightly scattered data in Problem 7.

From choosing a direction to finding one automatically.

In Problems 5–8, we *chose* two directions by looking at the scatterplot and then compared how much variance each captured. But in a dataset with 100 features, we cannot simply draw the data and guess the best direction. We need a way to find it from the data itself.

Question to discuss: Could a matrix that summarizes how features vary together help us find the best direction?

(9) From variance to the covariance matrix.

Return to our original dataset X from Problem 5. Each column has mean zero. Recall that for centered features x and y ,

$$\text{Var}(x) = \frac{1}{n-1} \sum_{i=1}^n x_i^2, \quad \text{Cov}(x, y) = \frac{1}{n-1} \sum_{i=1}^n x_i y_i.$$

(a) Calculate $\text{Var}(x)$, $\text{Var}(y)$, and $\text{Cov}(x, y)$ for the two columns of X . What does the positive covariance tell you?

(b) Assemble these quantities into the *covariance matrix*:

$$C = \begin{bmatrix} \text{Var}(x) & \text{Cov}(x, y) \\ \text{Cov}(x, y) & \text{Var}(y) \end{bmatrix}.$$

(c) Now compute $X^T X$.

(d) Explain what the entries of $X^T X$ are adding up to? What do you notice about $X^T X$? What happens if we divide each entry by $n - 1$? What might we name the new matrix C ?

$$\boxed{C = \frac{1}{n-1} X^T X} \quad (n = 4).$$

(10) How does the covariance matrix help us find directions?

For a unit direction $\mathbf{v}$, our projected data are the four numbers in $\mathbf{z} = X\mathbf{v}$. Since X is centered, the projected values are centered too.

(a) Recall that the sample variance of centered values $\mathbf{z}$ is $\text{Var}(\mathbf{z}) = \frac{1}{n-1} \mathbf{z}^T \mathbf{z}$. Substitute $\mathbf{z} = X\mathbf{v}$ and use $(X\mathbf{v})^T = \mathbf{v}^T X^T$ to complete:

$$\text{Var}(X\mathbf{v}) = \underline{\hspace{2cm}} = \underline{\hspace{2cm}} = \boxed{\mathbf{v}^T C \mathbf{v}}.$$

(b) We want the unit direction with the *greatest* projected variance. Suppose we find a unit eigenvector of C satisfying $C\mathbf{v} = \lambda\mathbf{v}$. Substitute this into the expression above:

$$\text{Var}(X\mathbf{v}) = \mathbf{v}^T C \mathbf{v} = \mathbf{v}^T (\lambda \mathbf{v}) = \lambda \underbrace{\mathbf{v}^T \mathbf{v}}_{?} = \underline{\hspace{2cm}}.$$

Why is it useful that $\mathbf{v}$ has length 1?

(c) Use $\det(C - \lambda I) = 0$ to find the two eigenvalues of the covariance matrix from Problem 9. Solve algebraically.

(d) Find a unit eigenvector for each eigenvalue. Compare the directions with the ones you chose geometrically in Problem 5.

(e) Compare the eigenvalues with the projected variances from Problem 8(a)–(b). What do you notice? Which eigenvector should we keep if we only want one feature?

(11) What if both directions have some variation?

Suppose a different centered dataset has covariance matrix

$$C_{\text{new}} = \begin{bmatrix} 2 & 1.8 \\ 1.8 & 2 \end{bmatrix},$$

with eigenvalues $\lambda_1 = 3.8$ and $\lambda_2 = 0.2$.

(a) What fraction of the total variance is represented by the first principal component? Show your calculation.

(b) What fraction would be discarded by keeping only the first component?

(c) Is keeping the direction with greater variance always the right decision for a *prediction* problem? Why or why not?

(12) Connect the ideas.

Complete the statements in your own words.

(a) The first principal component is the unit direction that _____.

(b) PCA uses perpendicular component directions because _____.

(c) When C is a covariance matrix, its eigenvalues represent _____.

(d) To transform a centered data matrix X into k principal-component coordinates, we use $Z =$ _____, where the columns of V_k contain _____.

Exit question: In one or two sentences, explain what PCA accomplishes and why eigenvectors help us do it.