Day 6: Eigenvalues, Eigenvectors, Diagonalization & PCA

Goal. Build geometric and computational intuition for eigenvalues/eigenvectors; understand diagonalization and why symmetric matrices are special; connect to PCA and stability in ML.

1. Core Definitions

Definition 1 (Eigenpair)**.** *For $A \in \mathbb{R}^{n\times n}$, a scalar $\lambda \in \mathbb{R}$ and nonzero vector $v \in \mathbb{R}^n$ form an* eigenpair *if*

$$Av = \lambda v.$$

Here λ is an eigenvalue *and v an* eigenvector.

Definition 2 (Characteristic Polynomial)**.** *Eigenvalues are the roots of $p_A(\lambda) = \det(A - \lambda I)$.*

2. Geometry

A scales v by λ without changing its direction. For general vectors x, Ax can rotate, shear, scale; eigenvectors are the special directions where only scaling occurs.

3. Diagonalization

If A has n linearly independent eigenvectors, stack them as columns of $P = [v_1 \ \cdots \ v_n]$ and let $D = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then

$$A = PDP^{-1}, \qquad A^k = PD^kP^{-1}.$$

This simplifies powers of A and dynamics $x_{t+1} = Ax_t$.

Proposition 1 (Spectral Theorem (real symmetric case))**.** *If $A = A^\top$, then A is diagonalizable by an orthogonal matrix: there exists Q with $Q^\top Q = I$ and $A = Q\Lambda Q^\top$ where Λ is diagonal and* real. *Eigenvectors can be chosen orthonormal.*

Remark 1 (Defective matrices)**.** *Non-symmetric A may be* non-diagonalizable *(Jordan blocks). Symmetric matrices avoid this pathology.*

4. Worked Examples (2×2)

Example A (diagonalizable)

$$A = \begin{bmatrix} 3 & 1 \\ 0 & 2 \end{bmatrix}, \quad p_A(\lambda) = (3-\lambda)(2-\lambda).$$

Eigenvalues: $\lambda_1 = 3,\ \lambda_2 = 2$.

For $\lambda_1 = 3$: $(A - 3I)v = 0 \Rightarrow \begin{bmatrix} 0 & 1 \\ 0 & -1 \end{bmatrix} v = 0 \Rightarrow v_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$.

For $\lambda_2 = 2$: $(A - 2I)v = 0 \Rightarrow \begin{bmatrix} 1 & 1 \\ 0 & 0 \end{bmatrix} v = 0 \Rightarrow v_2 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$.

Thus $P = \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix}$, $D = \mathrm{diag}(3, 2)$ and $A = PDP^{-1}$.

Example B (symmetric & orthogonal eigenvectors)

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

Characteristic polynomial: $(2-\lambda)^2 - 1 = \lambda^2 - 4\lambda + 3 = (\lambda-1)(\lambda-3)$, so $\lambda_1 = 3,\ \lambda_2 = 1$.

Eigenvectors (orthonormalized):

$$v_1 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \quad v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 \\ -1 \end{bmatrix}, \quad S = Q\Lambda Q^\top.$$

5. Rayleigh Quotient & Extremal Properties

For symmetric A and nonzero x,

$$R_A(x) = \frac{x^\top A x}{x^\top x}.$$

Proposition 2. *If $A = A^\top$ with eigenvalues $\lambda_{\min} \le \cdots \le \lambda_{\max}$, then $\lambda_{\min} \le R_A(x) \le \lambda_{\max}$, with equality iff x is an eigenvector for the extremal eigenvalue.*

This underlies PCA and power iteration.

6. Power Iteration (largest eigenvalue/eigenvector)

Given symmetric A and random $x_0 \ne 0$,

1. For $t = 0, 1, 2, \ldots$: $y_{t+1} = Ax_t$, then $x_{t+1} = y_{t+1}/\|y_{t+1}\|$.
2. Then $x_t \to v_{\max}$ (eigenvector), and $R_A(x_t) \to \lambda_{\max}$ if $\lambda_{\max}$ is unique and dominant.

7. PCA via Eigen-Decomposition

Given zero-mean data matrix $X \in \mathbb{R}^{n \times d}$ (rows are samples), the sample covariance is

$$\Sigma = \frac{1}{n} X^\top X \quad (\Sigma = \Sigma^\top,\ \Sigma \succeq 0).$$

- Eigenvectors of Σ are *principal directions*.
- Eigenvalues give *explained variance* along each direction.
- Project data onto top-k eigenvectors to get a k-dimensional representation with maximal retained variance.

8. Why This Matters in AI

- **Dimensionality Reduction (PCA):** Speeds training, reduces noise, and improves generalization by keeping directions of highest variance.
- **Stability & Dynamics:** The spectral radius $\rho(A) = \max_i |\lambda_i|$ governs stability of linear recurrences ($x_{t+1} = Ax_t$); in RNNs/Jacobians, eigenvalues > 1 can cause exploding activations/gradients, < 1 can cause vanishing.
- **Quadratic Forms:** Loss curvature ($x^\top Hx$) is controlled by Hessian eigenvalues; conditioning (ratio $\lambda_{\max}/\lambda_{\min}$) affects optimization speed.
- **Graph ML:** Graph Laplacian eigenvectors define smoothness bases; spectrum influences diffusion, clustering, and GNN behavior.
- **Whitening/Decorrelation:** Eigen-decomposition of covariance enables whitening transforms that standardize and decorrelate features.

9. Practical Tips

- Center data before PCA; optionally scale features to unit variance.
- For large d, use randomized SVD / iterative methods (power iteration, Lanczos) for top components.
- Watch conditioning: nearly equal leading eigenvalues slow convergence of power methods.
- For non-symmetric data operators, prefer SVD; for symmetric PSD matrices (covariance, kernel), eigen-decomposition is efficient and interpretable.

10. Mini-Exercises

1. Compute eigenvalues/eigenvectors of $\begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$ and relate to anisotropic scaling.
2. Show that for centered data, the principal component maximizes the Rayleigh quotient of Σ.
3. Implement power iteration and test it on the matrix S from Example B; compare the limit to the analytical eigenvector.
4. Prove that for symmetric A, eigenvectors corresponding to distinct eigenvalues are orthogonal.

Next

Day 7: Singular Value Decomposition (SVD), low-rank structure, and applications to recommender systems and compression.