

Principal Component Analysis (PCA)

1-Page Beginner Notes

What is PCA?

PCA is a dimensionality reduction technique that transforms data into a new coordinate system called principal components to capture maximum variance.

Why PCA is Needed

It reduces dimensionality, removes redundancy, improves visualization, and helps in noise reduction.

Mean-Centered Data

PCA works on mean-centered data. Variance is measured around the mean, and eigenvectors originate from the mean.

Eigenvectors (Principal Components)

Eigenvectors represent directions of maximum variance. They are orthogonal and ordered by importance. PC1 captures the highest variance.

Eigenvalues

Eigenvalues represent the amount of variance captured. Larger eigenvalue means a more important component.

PCA as Geometry

PCA is a rotation of axes. Data is projected onto new axes and low-variance directions are removed.

Visualization Insight

Data points show original data, the mean is shown as a central point, and arrows represent principal component directions.

Key Learning Points

Eigenvectors = directions, Eigenvalues = variance, PCA components are orthogonal, PCA starts from the mean.

One-Line Summary

PCA finds new orthogonal axes that capture maximum variance around the data mean for dimensionality reduction.