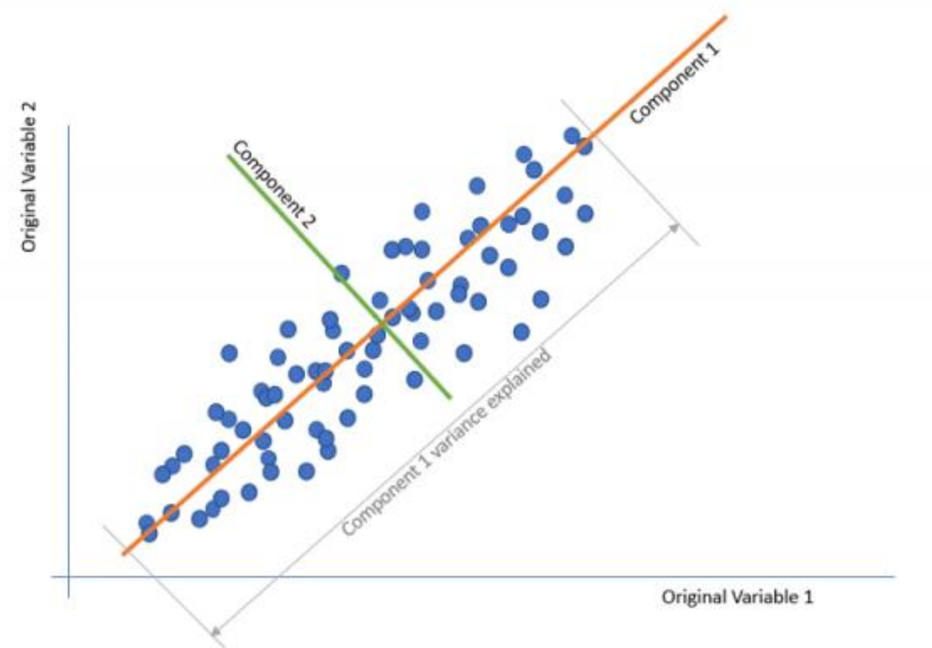
Exercise 5 Dimensionality Reduction Introduction

- PCA and LDA Basic Concepts and Exercise Overview

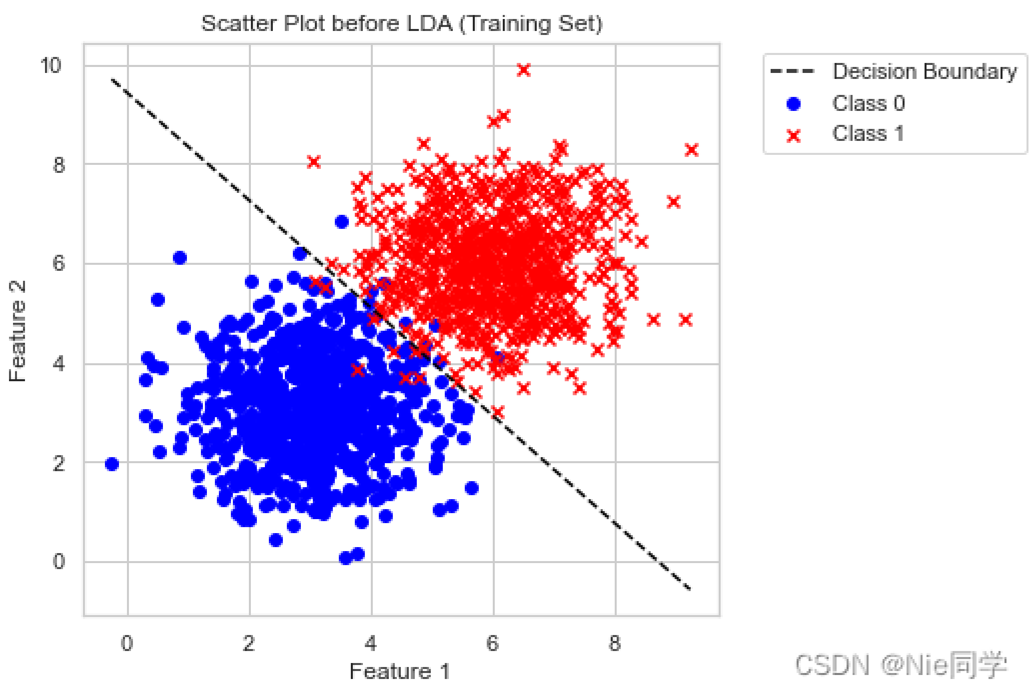
PCA Basic Introduction

- 📌 Preserve the largest variance in the data
- 🛠️ Project high-dimensional data to low-dimensional space
- ⚡ Select principal components via eigen decomposition of covariance matrix



LDA Basic Introduction

- Maximize the distance between classes
- Minimize the variance within classes
- Find the optimal projection direction



Problem 1 - (1) General Form of PCA Mapping

- ☑ Compute the mean of the data
- ☑ Centralize the data
- ☑ Compute the covariance matrix S
- ☑ Perform eigen decomposition
- Answer: Write down the eigenvalues and their corresponding eigenvectors.

$$\boldsymbol{\mu} = \frac{1}{n}\sum_{i=1}^{n}\mathbf{x}^{(i)}, \quad \mathbf{S} = \frac{1}{n}\sum_{i=1}^{n}(\mathbf{x}^{(i)} - \boldsymbol{\mu})(\mathbf{x}^{(i)} - \boldsymbol{\mu})^T$$

Problem 1

PCA is an unsupervised method of data dimensionality reduction.

- Write down the general form of data mapping from high to low-dimensionality space according to PCA. Explain the components of the mapping and comment on their orthogonality.
- Compute the principal components of the dataset:

$$\mathbf{X} = \{(1,2),(3,3),(3,5),(5,4),(5,6),(6,5),(8,7),(9,8)\}$$

Important: eigenvectors should be orthogonal

Problem 2 - (a) Between-class and Within-class Covariance

- Write down the formula for computing the within-class scatter matrix S_w
- Write down the formula for computing the between-class scatter matrix S_B
- **Note:** *You only need to write the formulas, no need to compute them*.

Problem 2

Let $p(x|\omega_i)$ be arbitrary densities with means μ_i and covariance matrices Σ_i, not necessarily normal for $i = 1, 2$. Let $y = w^T x$ be a projection, and let the induced one-dimensional densities $p(y|\omega_i)$ have means μ_i and variances σ_i.

(a) Write down the formulas for the between-class and inter-class covariance matrices.

Problem 2 - (b) Show Maximization of Criterion

- This question asks us to prove that the **optimal projection direction for Fisher discriminant** is given by:

$$\omega = ((\Sigma 1 + \Sigma 2)^{-1} (\mu 1 - \mu 2))$$

- The core concept is:
Maximizing the distance between class centers (after projection) divided by the total within-class variance, this ratio is called the **Fisher criterion**:

$$J(\omega) = \frac{(\omega^T (\mu_1 - \mu_2))^2}{\omega^T (\Sigma_1 + \Sigma_2) \omega}$$

- Write out the proof process
- Hints
  - This is a Rayleigh quotient, and its maximum value occurs at $\frac{\partial J(w)}{\partial w} = 0$
  - We only care about the direction of the projection, so scalar terms (e.g., $w^T S_w w$) can be ignored after taking derivatives.