Linear Discriminant Analysis

Machine Learning

IIIy CSE IDP and IDDMP

Refer : Stephen Marsland

Motivation

- Real-world classes often overlap in feature space
- Need to find a projection that best separates classes
- Reduce dimensionality while preserving class discrimination
- Maximizes distance between class means
- Minimizes class-internal scatter

What is LDA?

- Linear Discriminant Analysis (LDA) is a supervised dimensionality-reduction technique
- Seeks a linear combination of features that best separates two or more classes
- Projects data onto a lower-dimensional subspace by maximizing class separability
- Commonly used for feature extraction, classification preprocessing

Within-Class Scatter

- Define class means: $\mu_1, \mu_2, \ldots, \mu_k$ and overall mean μ
- For each class c with probability p_c, compute covariance of its samples around μ_c
- Sum weighted covariances to get within-class scatter:

$$S_W = \sum_{\text{classes } c} \sum_{j \in c} p_c (\mathbf{x}_j - \boldsymbol{\mu}_c)(\mathbf{x}_j - \boldsymbol{\mu}_c)^T.$$

- Measures how tightly samples cluster around their class means; Lower $S_W \Rightarrow$ better class compactness

Between-Class Scatter

- Tight within-class clusters are desirable (small S_W), but classes must also be well separated

- Between-class scatter measures distance of each class mean from the overall mean

- Defined as: $$S_B = \sum_{\text{classes } c} (\boldsymbol{\mu}_c - \boldsymbol{\mu})(\boldsymbol{\mu}_c - \boldsymbol{\mu})^T.$$

- Large $S_B \Rightarrow$ class centers are far apart

Graph depicting S_W and S_B

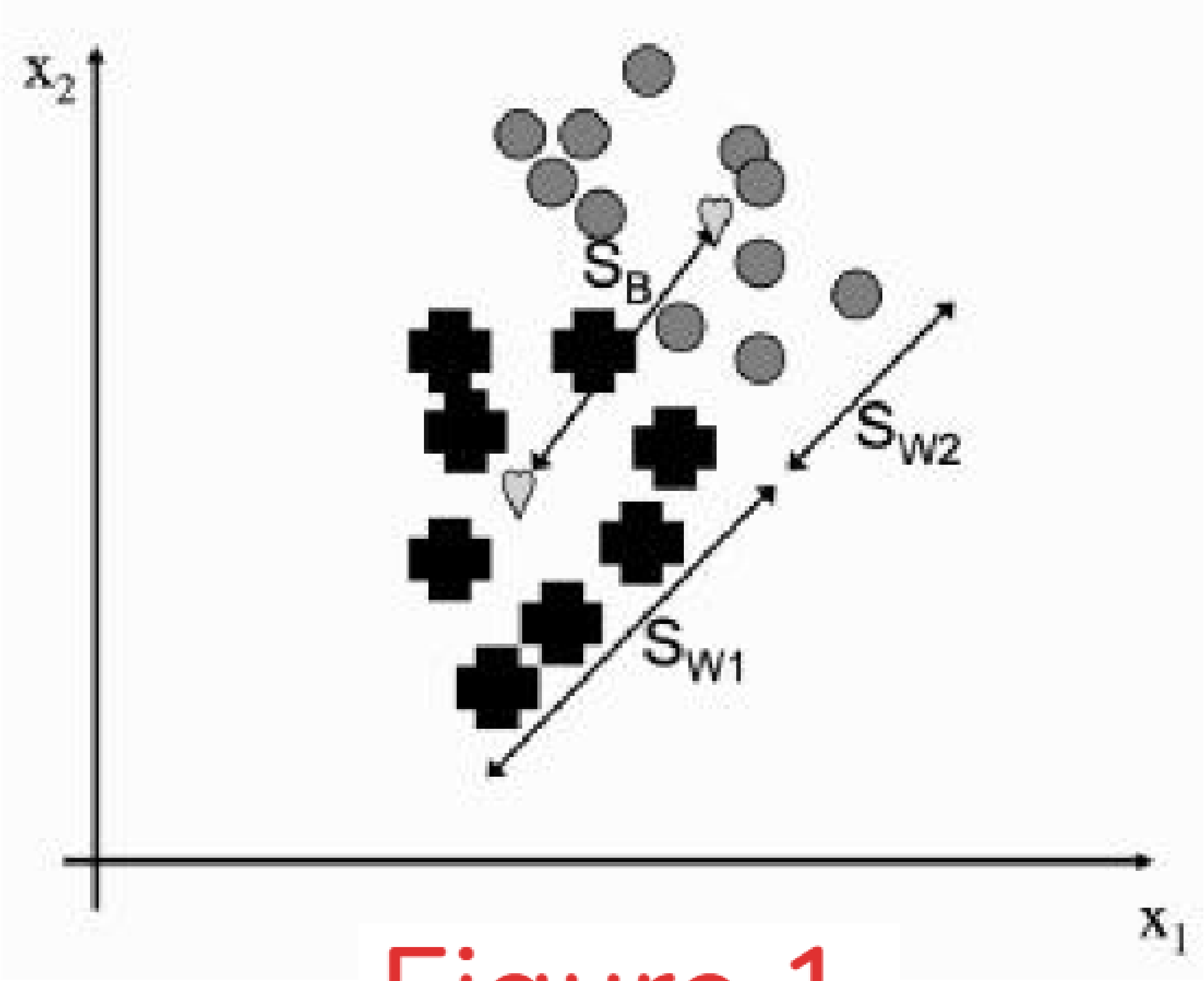

Figure 1

Interpreting Scatter Ratios & Projections

- **Visualizing S_W vs. S_B**: Figure 1 illustrates how within-class scatter (S_W) and between-class scatter (S_B) relate to class compactness and separation

- **Separation Criterion**: “Good” datasets have a large ratio S_B/S_W , i.e., tightly clustered classes that are far apart

- **Dimensionality Reduction Goal**: When projecting into fewer dimensions, we still want S_B/S_W to remain as large as possible

Example Projections

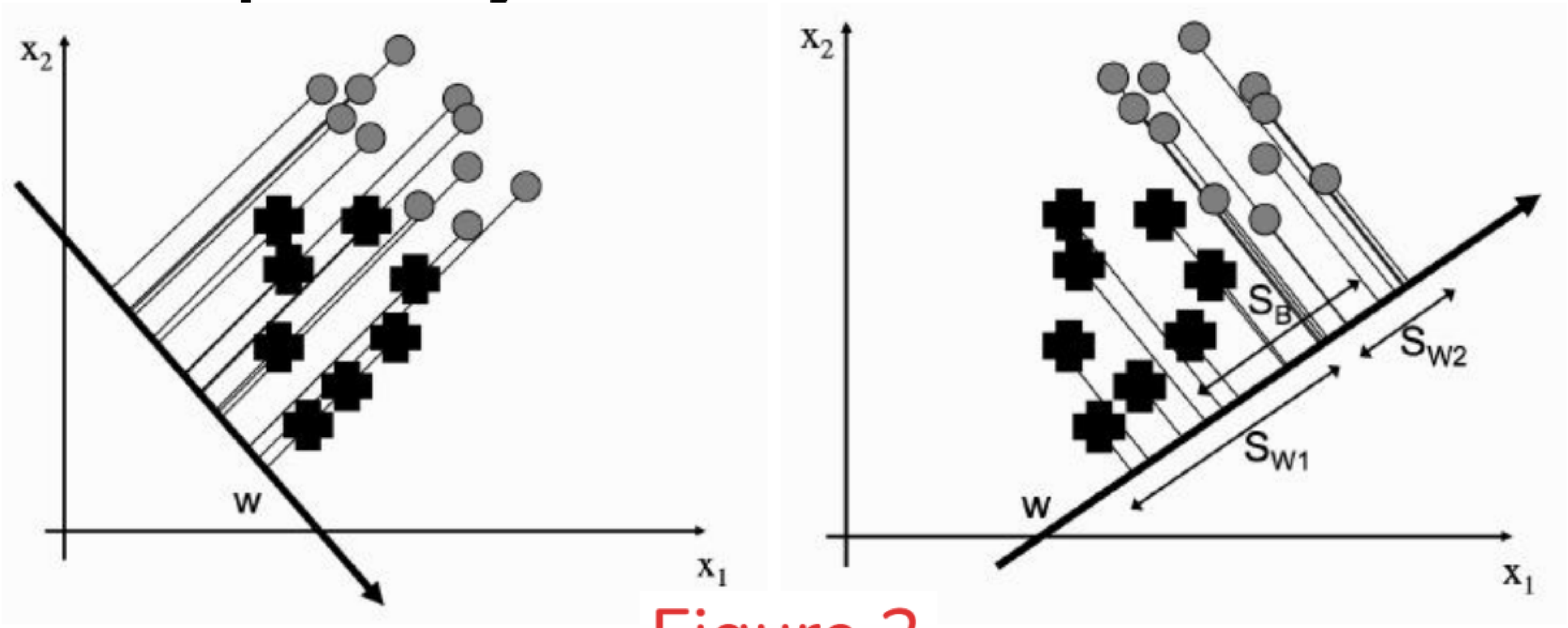

Figure 2

- **Left**: Low S_B/S_W, classes overlap after projection
- **Right**: High S_B/S_W, classes well separated after projection

Projection onto a Line

- Any line in feature space can be represented by a weight vector **w** (a row of the projection matrix **W**)

- Project a data point **x** onto **w** by: $\mathbf{z} = \mathbf{w}^T\mathbf{x}$
  Since, $\mathbf{w}^T\mathbf{x} = ||\mathbf{w}||\ ||\mathbf{x}|| \cos\theta$, setting $||\mathbf{w}|| = 1$ makes **z** the signed distance along **w**

- Every datapoint—and even class means—can be projected: $\mu'_c = \mathbf{w}^T \cdot \boldsymbol{\mu}_c.$

- After projection, we can analyze how the within- and between-class scatters transform in this 1D space

Deriving the Optimal Projection

- **Projection Substitution**
  - Replace each data point x_j with its projection $z_j = w^T x_j$ in the scatter definitions
  - Within-class scatter becomes $w^T S_W w$ (Eq. 1)
  - Between-class scatter becomes $w^T S_B w$ (Eq. 2)
- **Scatter Ratio**

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

We wish to maximize $J(w)$ with respect to w.

Deriving the Optimal Projection (cont.)

- **Stationary Condition**

  Differentiate $J(w)$ and set derivative to zero gives

$$S_B w\,(w^T S_W w) \;-\; S_W w\,(w^T S_B w) \;=\; 0 \qquad \text{(Eq. 3)}$$

- **Generalized Eigenvalue Problem**
  - Re-arrangement

$$S_W \mathbf{w} = \frac{\mathbf{w}^T S_W \mathbf{w}}{\mathbf{w}^T S_B \mathbf{w}} S_B \mathbf{w}. \qquad \text{(Eq. 4)}$$

  - Equivalently,

$$S_B w = \lambda\, S_W w \quad \Longrightarrow \quad S_W^{-1} S_B\, w = \lambda\, w$$

Solution

- **Optimal Directions**
  - Compute the eigenvectors of $S_W^{-1}S_B$
  - Choose the eigenvectors corresponding to the largest eigenvalues λ to maximize class separability
- **Two-Class Special Case**
  - For two classes, $S_B = (\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$
  - Then

$$S_W^{-1} S_B = S_W^{-1}(\mu_1 - \mu_2)(\mu_1 - \mu_2)^T$$

  - The dominant eigenvector points in the direction

$$w \propto S_W^{-1}(\mu_1 - \mu_2)$$

  - Scalar constants (e.g.\ ratios of p_c or total counts) do not change the direction

Summary

- **Goal of LDA**: Find a low-dimensional projection that maximizes class separability
- **Key Quantities**:
  - Within-class scatter S_W: measures compactness of each class
  - Between-class scatter S_B: measures separation of class means
- **Optimization Criterion**: Maximize the ratio

$$J(w) = \frac{w^T S_B w}{w^T S_W w}$$

- **Solution**:
  - Solve the generalized eigenvalue problem $S_W^{-1} S_B \, w = \lambda \, w$
  - Projection vectors w are the eigenvectors with the largest eigenvalues
- **Special Case (2 classes)**:

$$w \propto S_W^{-1}(\mu_1 - \mu_2)$$

- **Outcome**: Reduced-dimensional representation that preserves discriminative information for classification.