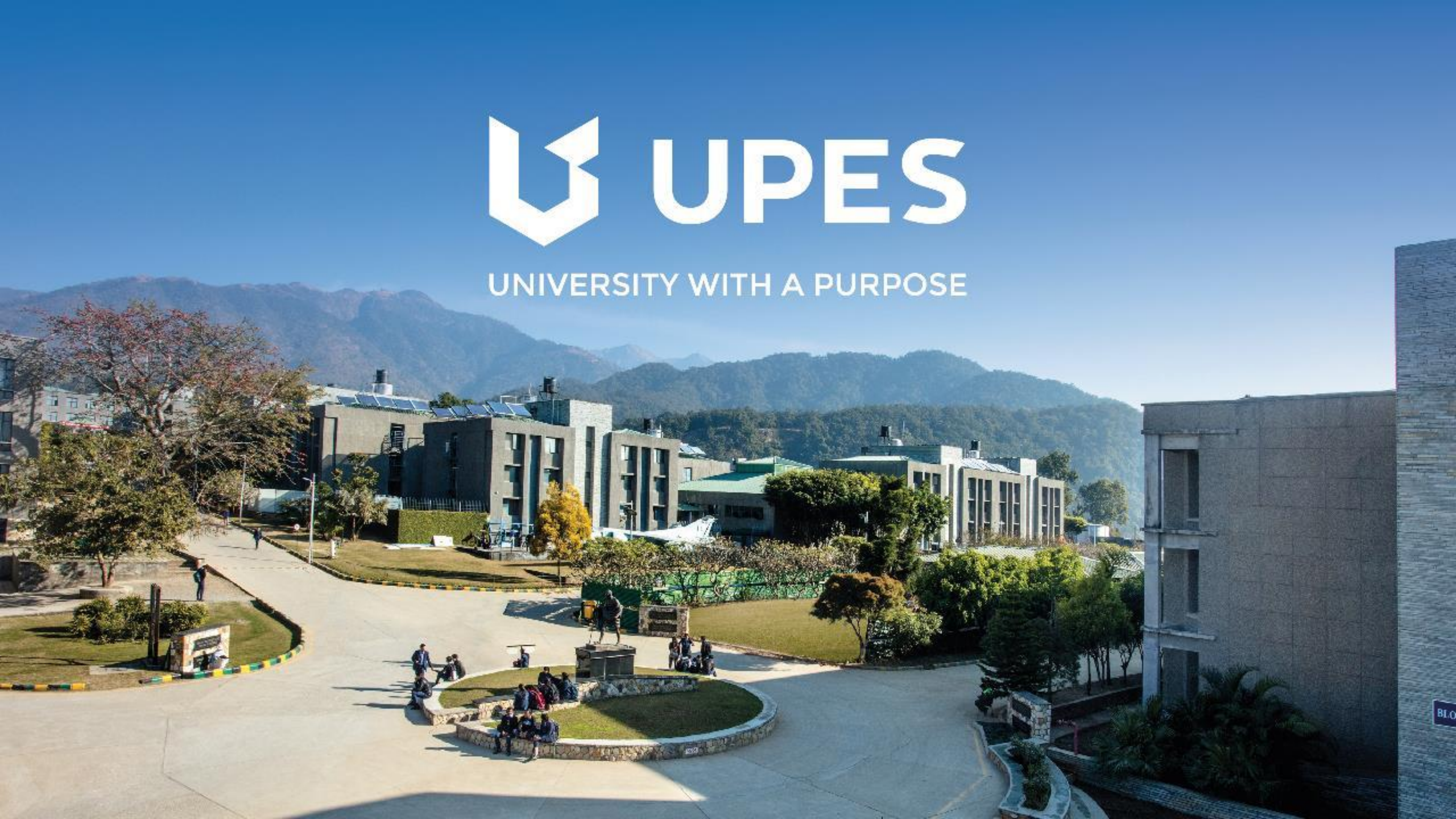
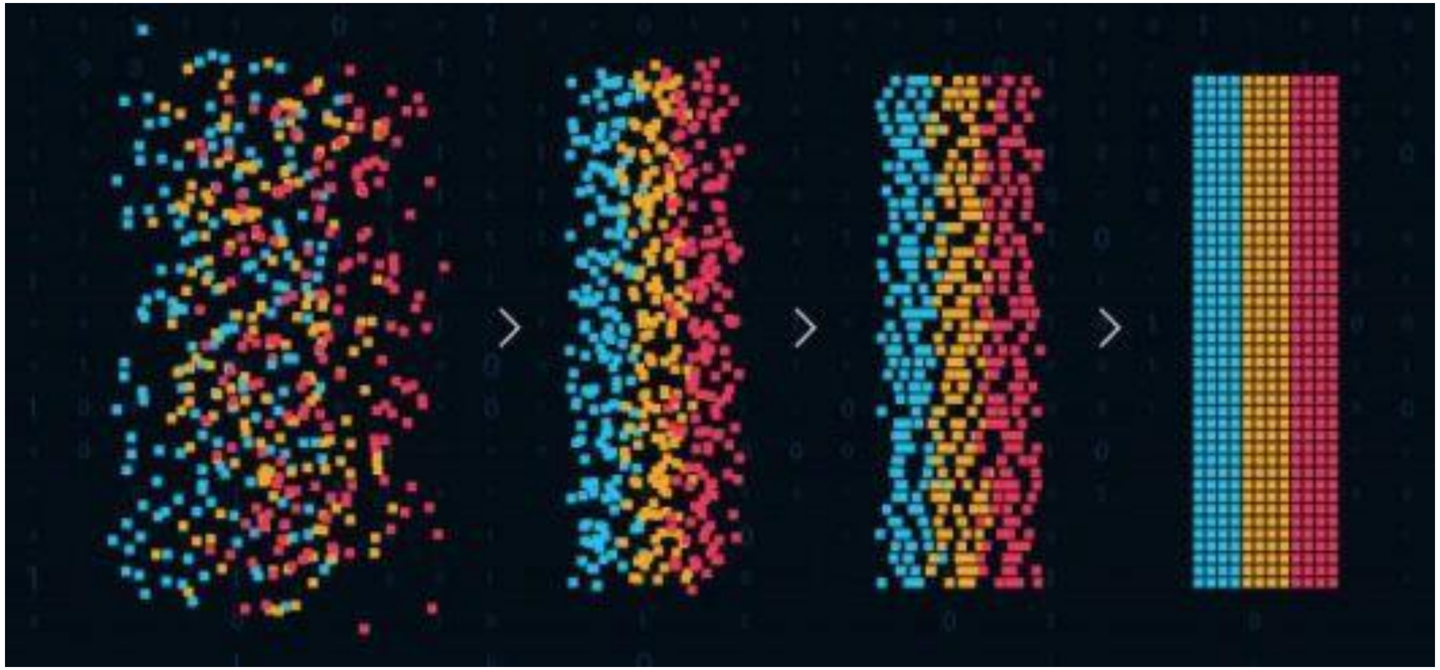

UPES
UNIVERSITY WITH A PURPOSE

Pattern and Anomaly Detection



Source: Edureka

B. Tech., CSE + AI/ML
Dr Gopal Singh Phartiyal
25/10/2021

Recap: Linear Models for Classification

- **Classification:** Assign input x to one of the class labels
- Why we call it linear?
- Linearly separarble dataset and Hyperplanes

- **Linear Discriminant Functions**
  - **Two class**
  - **Multiclass**
    - **3 approaches:** one-versus-the-rest, one-versus-one, and k classifiers
    - Decision region ambiguities

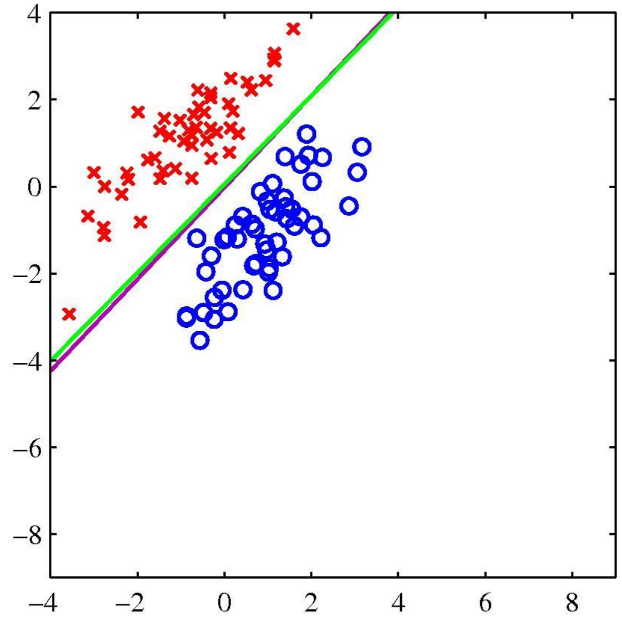
Li-Mod for Classification: Compute W

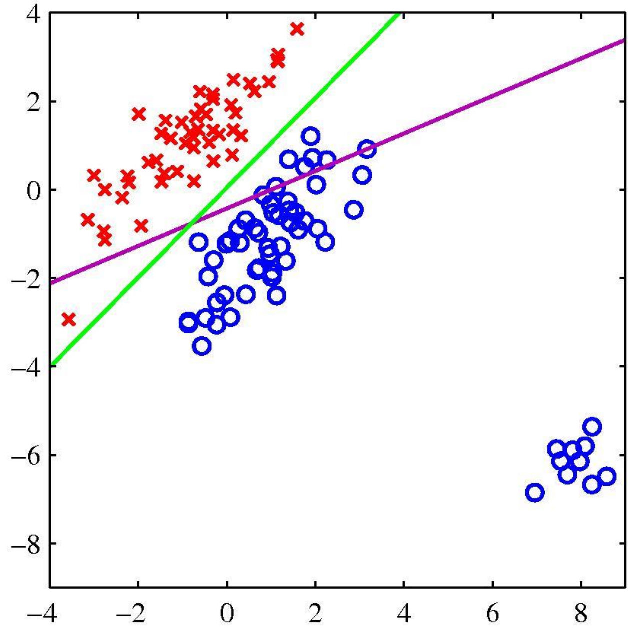
- First approach: Least squares
  - (tildeh when w_0 is included)

$$\widetilde{\mathbf{W}} = (\widetilde{\mathbf{X}}^{\mathrm{T}}\widetilde{\mathbf{X}})^{-1}\widetilde{\mathbf{X}}^{\mathrm{T}}\mathbf{T}$$

- Same issue with least squares

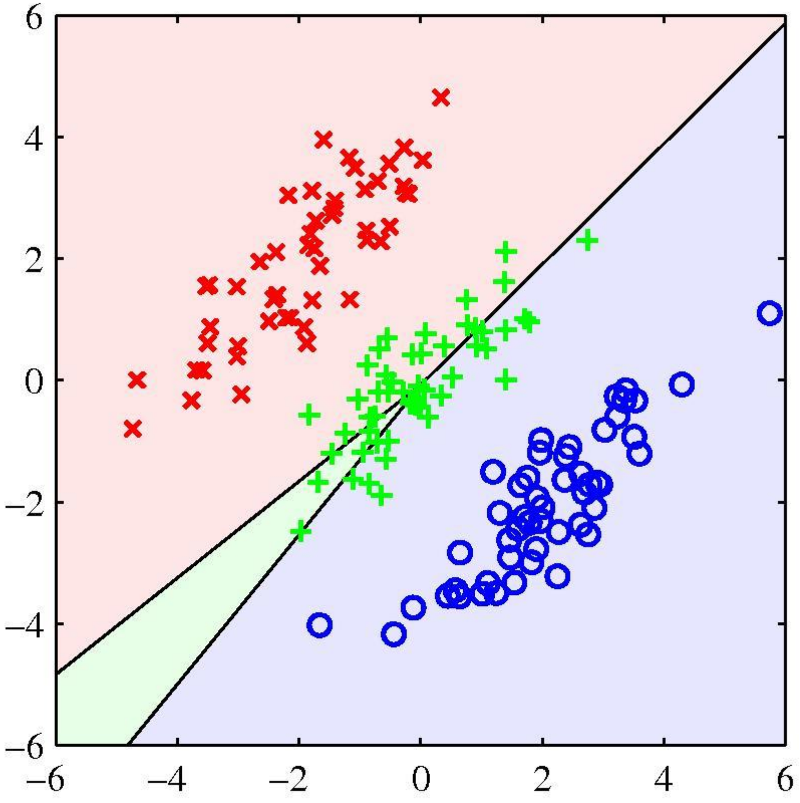
$$E_D(\widetilde{\mathbf{W}}) = \frac{1}{2}\mathrm{Tr}\left\{(\widetilde{\mathbf{X}}\widetilde{\mathbf{W}} - \mathbf{T})^{\mathrm{T}}(\widetilde{\mathbf{X}}\widetilde{\mathbf{W}} - \mathbf{T})\right\}$$



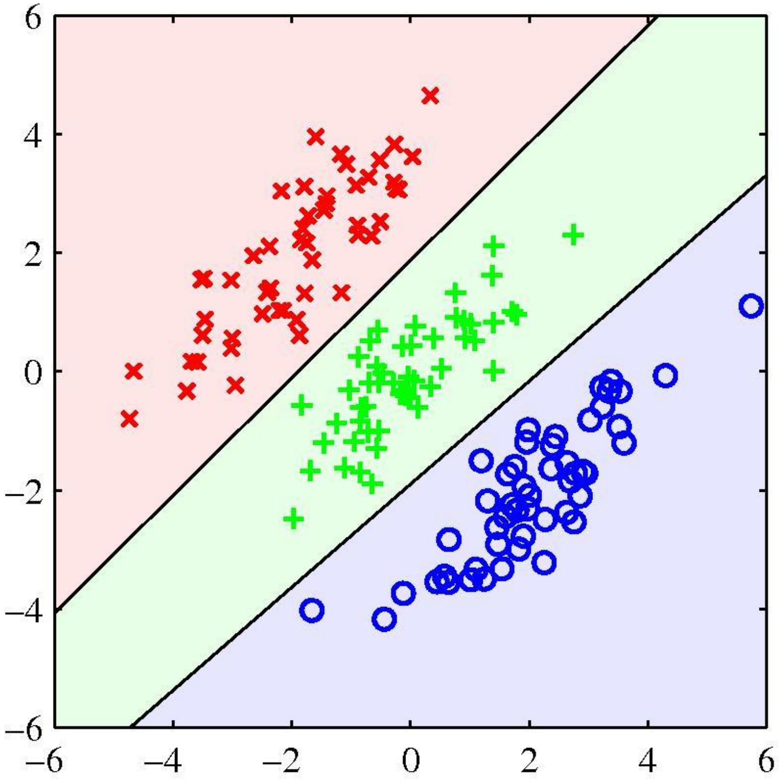


Li-Mod for Classification: Compute W

- Multiple class



- Least squares



- Logistic regression

Li-Mod for Classification: Fisher's Linear Discriminant

- **Concept**: Imagine linear classification models in terms of dimensionality reduction.

$$y = \mathbf{w}^{\mathrm{T}}\mathbf{x}.$$

$y \geqslant -w_0$ for C_1

otherwise C_2

- Loss of information in the process of projecting data into lower dimensions.
- Find a projection that **maximizes** the **class separation.**
- FLD is one such projection method (therefore also used in dimension reduction)

Li-Mod for Classification: Fisher's Linear Discriminant

- Consider two-class problem: N_1 points in C_1 and N_2 points in C_2
- Compute mean of both.

$$\mathbf{m}_1 = \frac{1}{N_1} \sum_{n \in \mathcal{C}_1} \mathbf{x}_n, \qquad \mathbf{m}_2 = \frac{1}{N_2} \sum_{n \in \mathcal{C}_2} \mathbf{x}_n.$$

- Maximize the difference in projected space i.e.

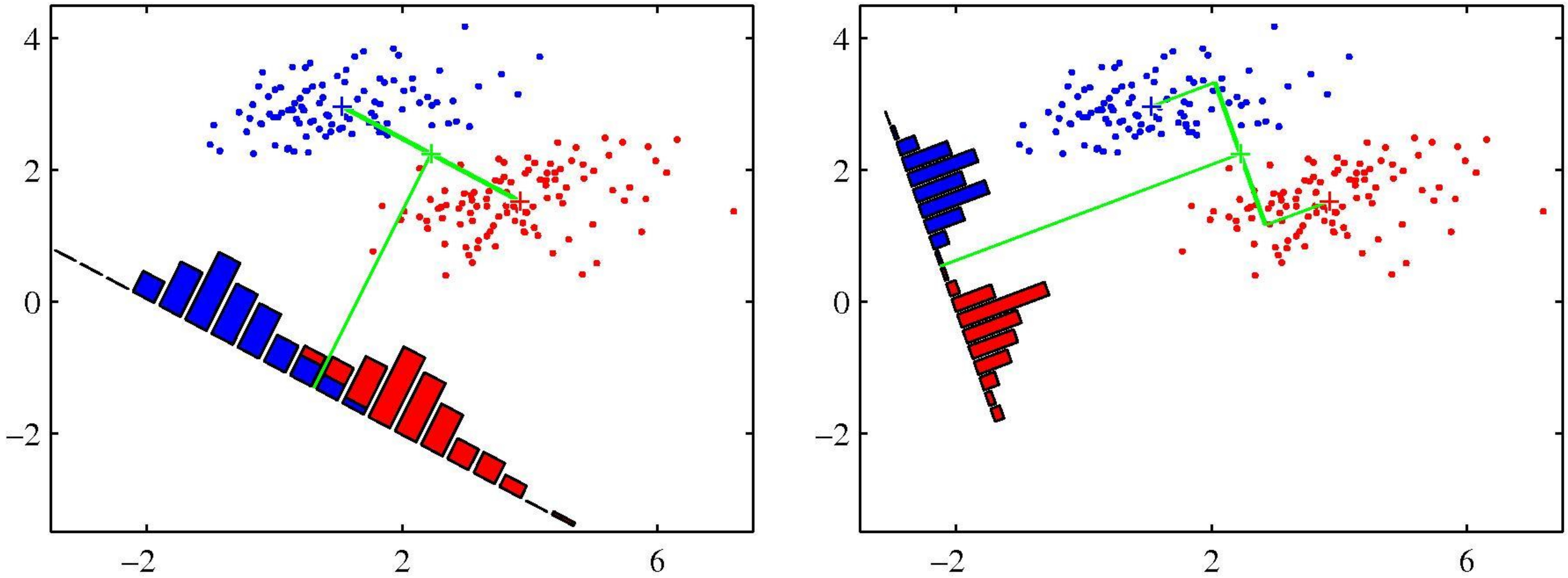
- Maximize

$$m_2 - m_1 = \mathbf{w}^{\mathrm{T}}(\mathbf{m}_2 - \mathbf{m}_1)$$

- The within class variance

$$s_k^2 = \sum_{n \in \mathcal{C}_k} (y_n - m_k)^2$$

Li-Mod for Classification: Fisher's Linear Discriminant



Fisher criterion $J(\mathbf{w}) = \frac{(m_2 - m_1)^2}{s_1^2 + s_2^2}$

$$\mathbf{w} \propto \mathbf{S}_{\mathrm{W}}^{-1}(\mathbf{m}_2 - \mathbf{m}_1).$$

$$J(\mathbf{w}) = \frac{\mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{B}}\mathbf{w}}{\mathbf{w}^{\mathrm{T}}\mathbf{S}_{\mathrm{W}}\mathbf{w}}$$

FLD

Li-Mod for Classification: Fisher's Linear Discriminant

- Multi-class

$$\mathrm{S_W} = \sum_{k=1}^{K} \sum_{n \in \mathcal{C}_k} (\mathbf{y}_n - \boldsymbol{\mu}_k)(\mathbf{y}_n - \boldsymbol{\mu}_k)^{\mathrm{T}}$$

$$\mathrm{S_B} = \sum_{k=1}^{K} N_k (\boldsymbol{\mu}_k - \boldsymbol{\mu})(\boldsymbol{\mu}_k - \boldsymbol{\mu})^{\mathrm{T}}$$

$$= \frac{1}{N_k} \sum_{n \in \mathcal{C}_k} \mathbf{y}_n, \qquad \boldsymbol{\mu} = \frac{1}{N} \sum_{k=1}^{K} N_k \boldsymbol{\mu}_k.$$

$$J(\mathbf{w}) = \mathrm{Tr}\left\{ (\mathbf{W} \mathrm{S_W} \mathbf{W}^{\mathrm{T}})^{-1} (\mathbf{W} \mathrm{S_B} \mathbf{W}^{\mathrm{T}}) \right\}$$

Thank You

UPES

UNIVERSITY WITH A PURPOSE