E9 205 Machine Learning for Signal Processing

Supervised-Dimensionality-Reduction.
Decision Theory
Probability Distributions

26-08-2019

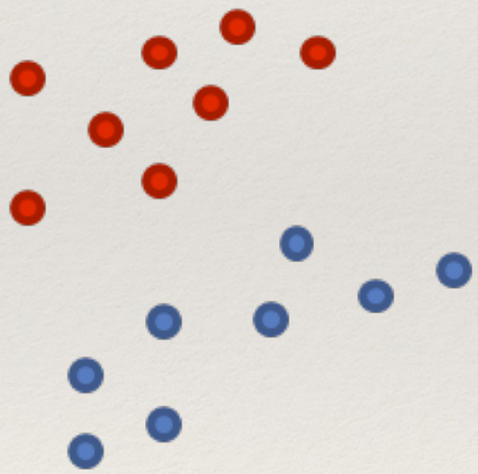
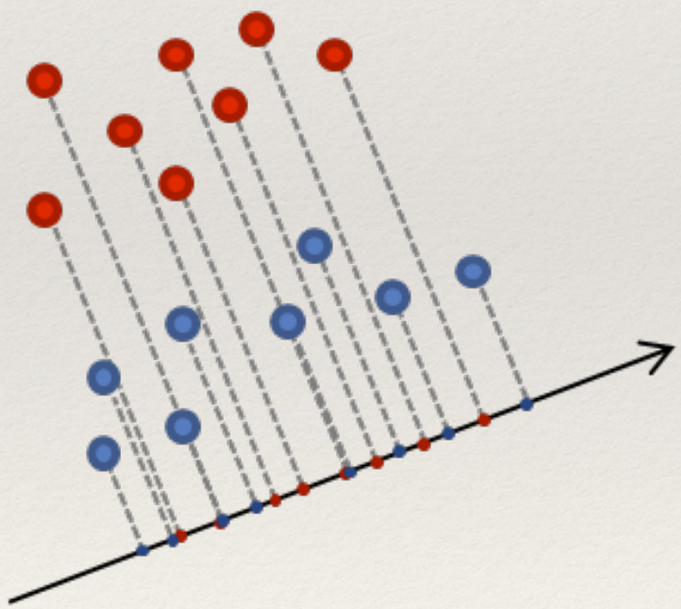
Established 1911

INDIAN INSTITUTE OF SCIENCE
भारतीय विज्ञान संस्थान

Advantages and Disadvantages of PCA

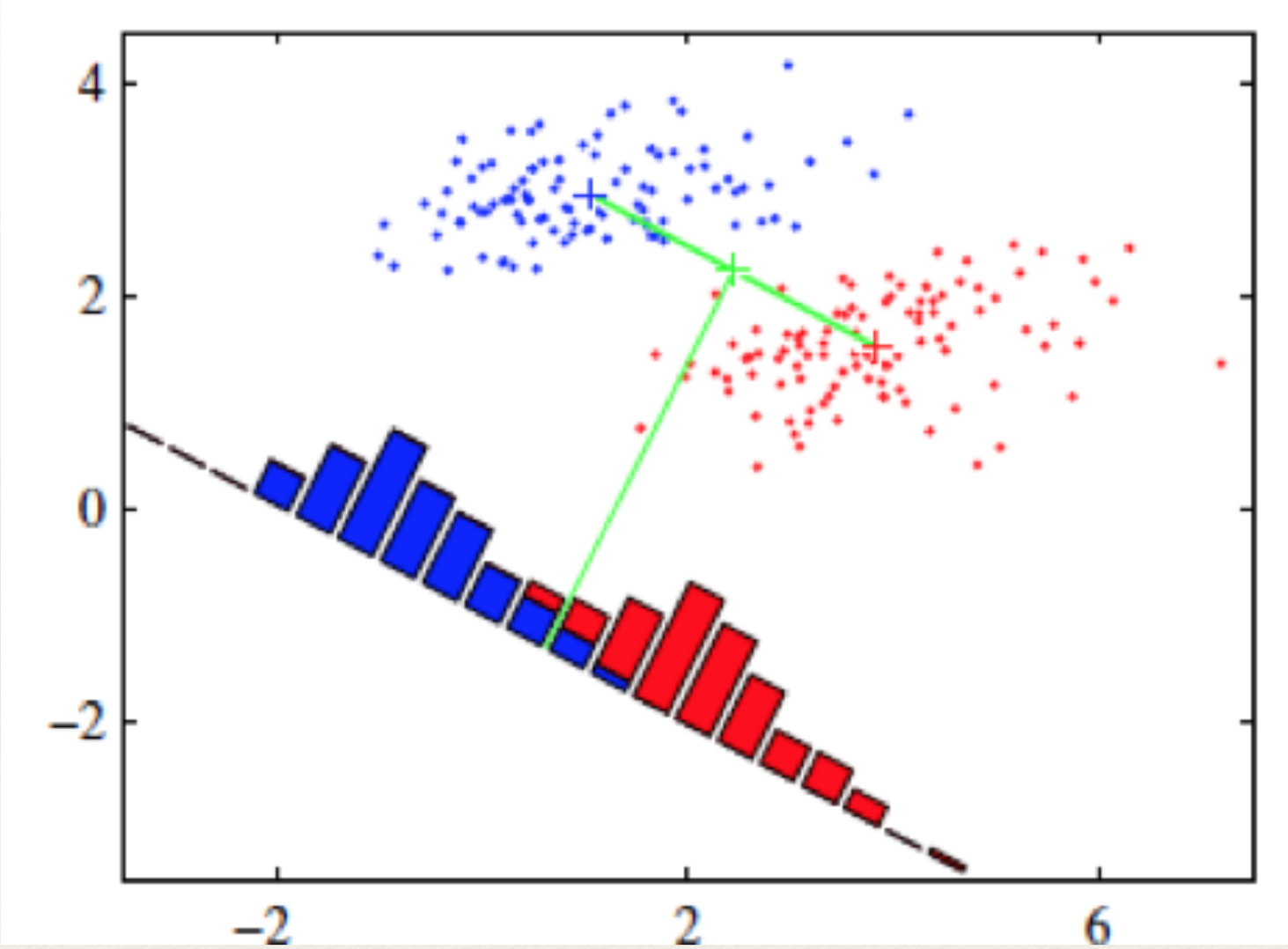
- Simple linear transform
  - Eigen decomposition of Data Covariance matrix is straight forward.
    - PCA for high dimensional data ?
- Variance maximization may not be the ideal loss function in dimensionality reduction.
- If the data contains discrete class labels, we can do better than PCA to maximize class separation.

Need for Supervised Dimensionality Reduction

Labelled data	PCA projection: Maximising the variance of the whole set
	

Linear Discriminant Analysis

Without the Within Class Factor



Linear Discriminant Analysis

Find a linear transform $f(\mathbf{x}) = \mathbf{w}^T\mathbf{x}$ with a criterion which maximizes the class separation

- Maximize the between class distance in the projected space while minimizing the within class covariance

$$J = \frac{\mathbf{w}^T \boldsymbol{S}_b \mathbf{w}}{\mathbf{w}^T \boldsymbol{S}_w \mathbf{w}}$$

$$\boldsymbol{S}_b = \sum_{k=1}^{K} N_k(\mathbf{m}_k - \boldsymbol{m})(\mathbf{m}_k - \boldsymbol{m})^T \quad \boldsymbol{S}_w = \sum_{k=1}^{K} \sum_{n \in C_k} (\mathbf{x}_n - \boldsymbol{m}_k)(\mathbf{x}_n - \boldsymbol{m}_k)^T$$

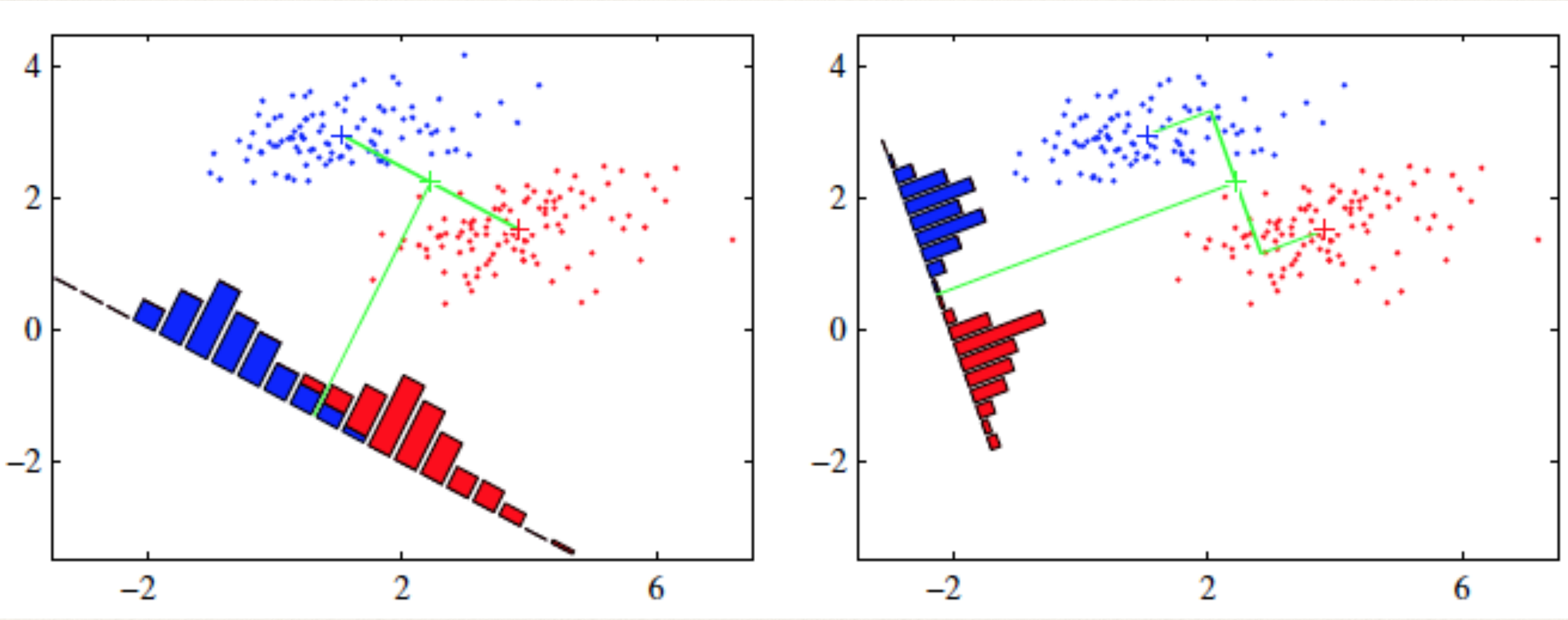
- Generalized Eigenvalue problem
- Eigenvectors of $\boldsymbol{S}_w^{-1}\boldsymbol{S}_b$

PRML - C. Bishop (Sec. 4.1.4, Sec. 4.1.6)

Linear Discriminant Analysis

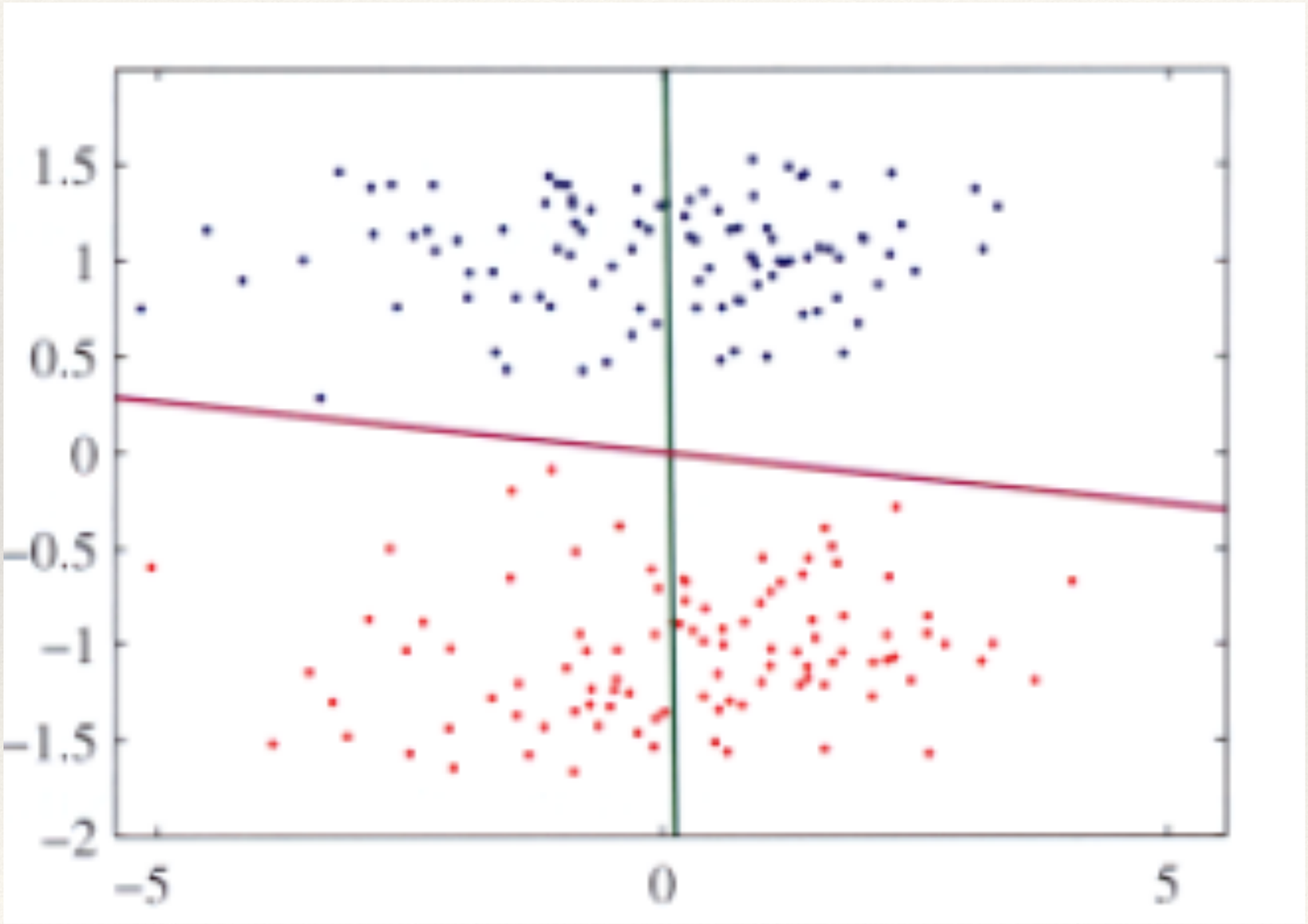
Projecting on line joining means

Fisher Discriminant



PRML - C. Bishop (Sec. 4.1.4, Sec. 4.1.6)

PCA versus LDA



PRML - C. Bishop (Sec. 4.1.4, Sec. 4.1.6)

PCA versus LDA

Labelled data	PCA projection: Maximising the variance of the whole set	LDA projection: Maximising the distance between groups