Supervised Classification

Supervised classification

Fisher's Linear Discriminant

Computational aspects

Supervised Classification

- Given a training set: $\{(x(t), y(t))\}_{t=1}^{m}$

  where $x(t) \in R^n, \quad y(t) \in \{-, +\}$
- Predict the class {-,+} for unseen inputs x.

- Example: medical diagnosis
  - x = (age, blood pressure, temperature, ...)^T
  - y = is it myocardial infarction (heart attack)?

Linear Discrimination

- Obtain the class as follows:

$$\begin{aligned} w^{\mathrm{T}}x > b &\quad \rightarrow \quad y = + \\ w^{\mathrm{T}}x \leq b &\quad \rightarrow \quad y = - \end{aligned}$$

- How do we find the weights w and the threshold b?
- Together <w,b> define a linear separator.
- Goal: “learn” an optimal separator for the given training data.

Plan Z

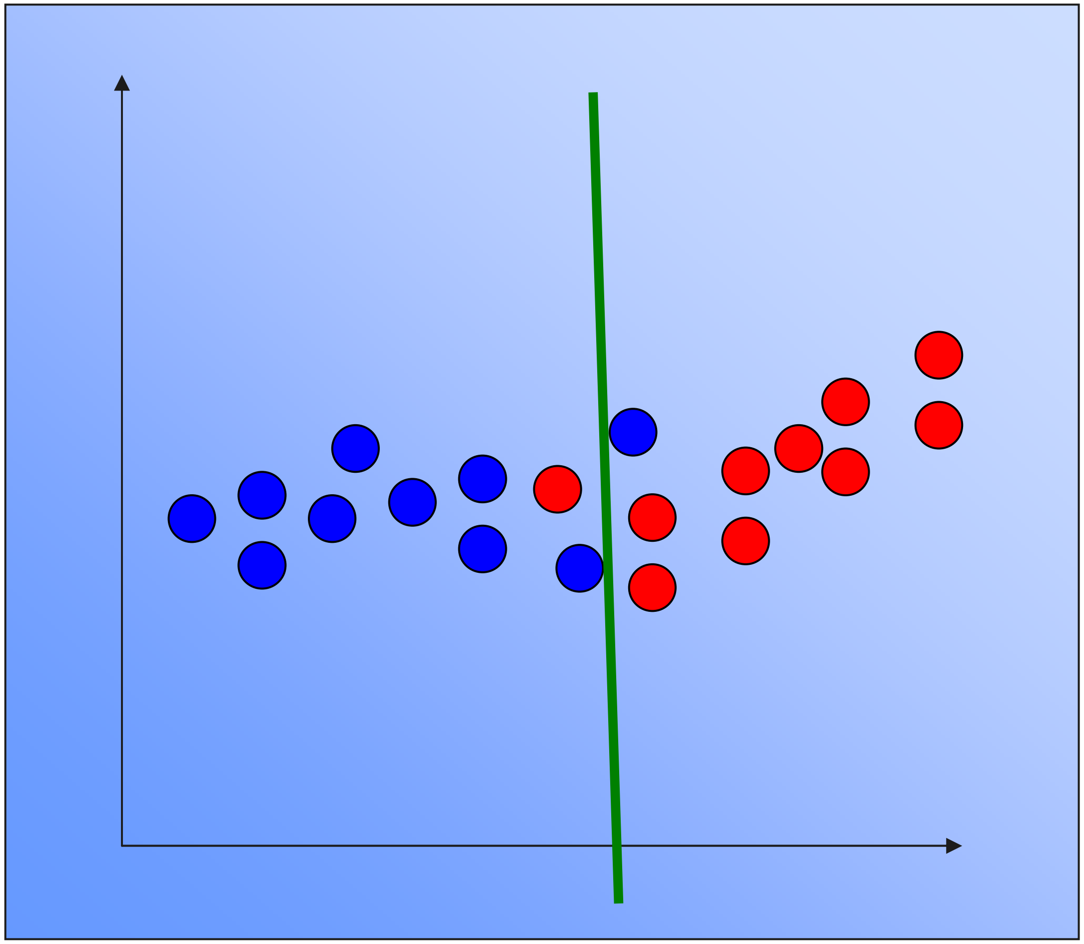
- Minimize the training error:

$$J(w,b) = \left|\left\{y=+, w^{\mathrm{T}}x \le b\right\}\right| + \left|\left\{y=-, w^{\mathrm{T}}x > b\right\}\right|$$

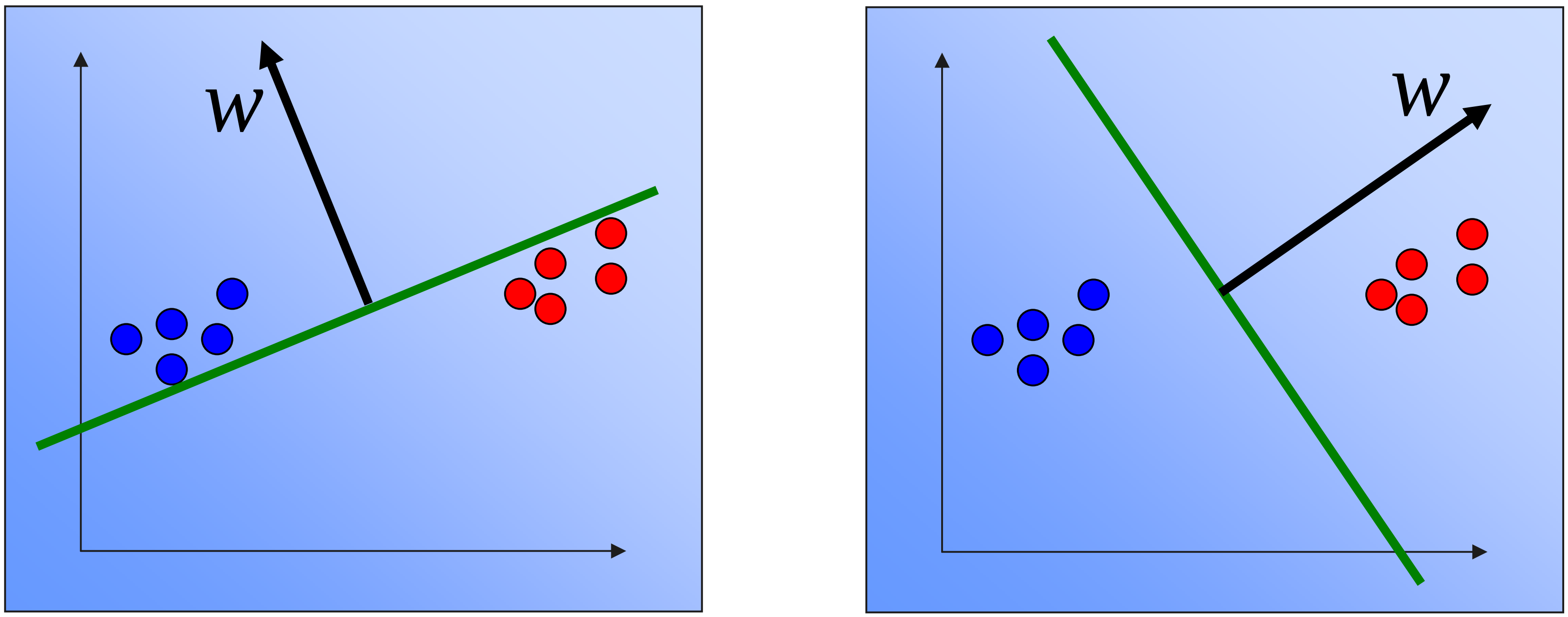
$$\langle w,b\rangle = \arg\min_{w,b} J(w,b)$$

- Difficult problem (NP-complete)
- Solution might not be unique
- In practice, not all errors are equal

Example (n = 2)

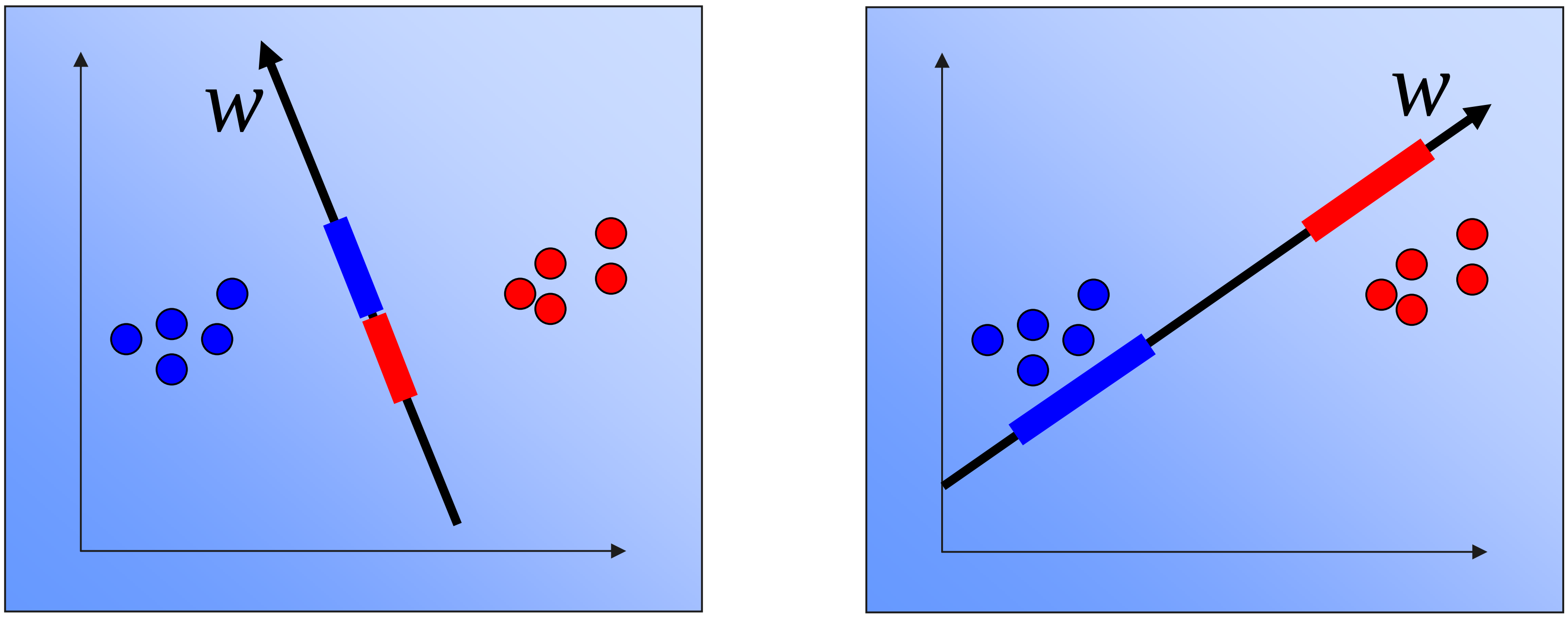


Example (n = 2)



$$w^{\mathrm{T}} x = b$$

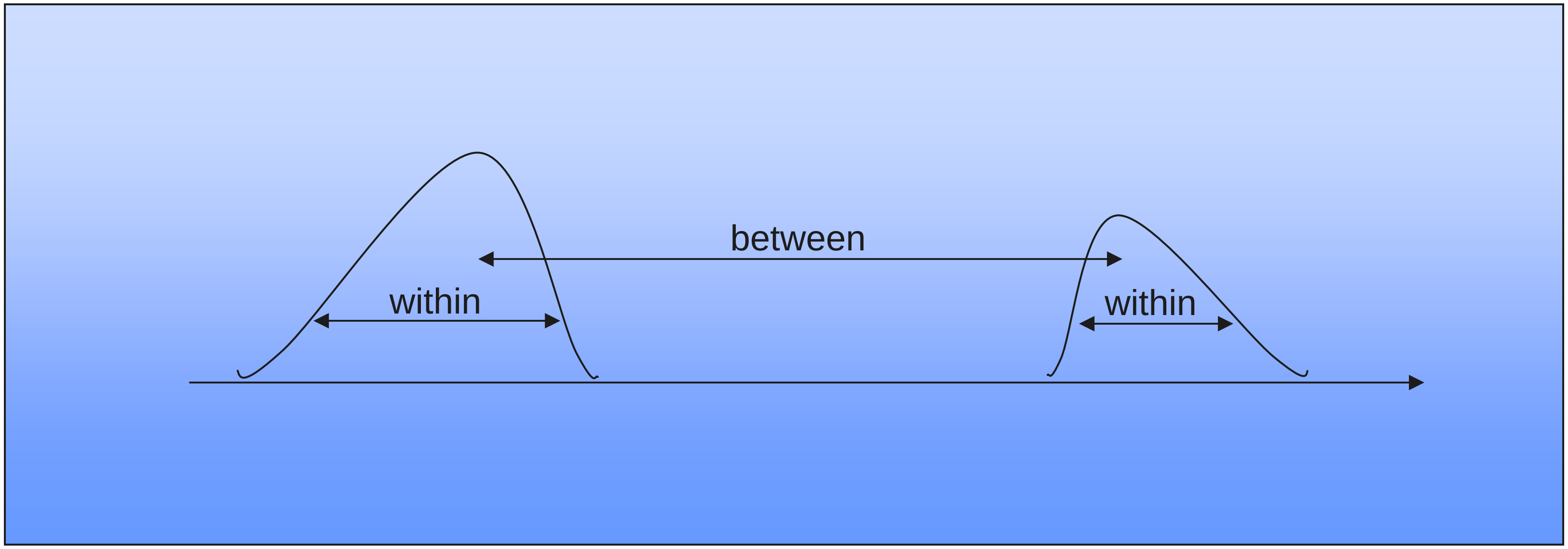
Example (n = 2)



$$w^{\mathrm{T}} x = b$$

Fisher’s Linear Discriminant (FLD)

- R. A. Fisher, “The use of multiple measurements in taxonomic problems” [1936].
- (Informal) Goal: maximize the ratio of the variance **between** the classes to the variance **within** the classes.



FLD 2

- Assume that the examples (projected onto the optimal direction w) form two distributions with means m_-, m_+, and variances σ_-^2, σ_+^2:

$$m_- = E\left(w^{\mathrm{T}}x \,|\, y = -\right) \qquad m_+ = E\left(w^{\mathrm{T}}x \,|\, y = +\right)$$

$$\sigma_-^2 = V\left(w^{\mathrm{T}}x \,|\, y = -\right) \qquad \sigma_+^2 = V\left(w^{\mathrm{T}}x \,|\, y = +\right)$$

- Solve for

$$w^* = \arg\max_w \frac{(m_- - m_+)^2}{\left(\sigma_-^2 + \sigma_+^2\right)}$$

FLD 3

- If x is a random variable with mean μ and covariance matrix Σ then:

$$E\left(w^{\mathrm{T}}x\right) = w^{\mathrm{T}}\mu$$

$$V\left(w^{\mathrm{T}}x\right) = w^{\mathrm{T}}\Sigma w$$

- Therefore:

$$m_{+} = E\left(w^{\mathrm{T}}x \,|\, y = +\right) = w^{\mathrm{T}}\mu_{+}$$

$$\sigma_{+}^{2} = V\left(w^{\mathrm{T}}x \,|\, y = +\right) = w^{\mathrm{T}}\Sigma_{+}w$$

FLD 4

- Let the two classes of observations have means μ_-, μ_+ and covariances Σ_-, Σ_+.
- Maximize

$$\frac{(m_- - m_+)^2}{(\sigma_-^2 + \sigma_+^2)} = \frac{\left(w^T\mu_- - w^T\mu_+\right)^2}{w^T\Sigma_- w + w^T\Sigma_+ w}$$

$$= \frac{w^T(\mu_- - \mu_+)(\mu_- - \mu_+)^T w}{w^T(\Sigma_- + \Sigma_+)w}$$

$$= \frac{w^T A w}{w^T B w}$$

Optimal Solution

- The solution to the optimization problem on the previous slide is given by the dominant generalized eigenvector of (A,B). This vector satisfies:

$$Ax = \lambda Bx$$

FLD – Summary

- Given a training set $\{(x(t), y(t))\}_{t=1}^{m}$
- Compute:

$$\mu_+ = \frac{1}{N_+} \sum_{t:y(t)=+} x(t)$$

$$\Sigma_+ = \frac{1}{N_+} \sum_{t:y(t)=+} (x(t) - \mu_+)(x(t) - \mu_+)^{\mathrm{T}}$$

$$A = (\mu_- - \mu_+)(\mu_- - \mu_+)^{\mathrm{T}} \qquad B = (\Sigma_- + \Sigma_+)$$

- Compute the dominant generalized eigenvector of (A,B): $Ax = \lambda Bx$

Generalized Eigenvectors

- One option is to convert to an ordinary eigenvector problem:

$$Ax = \lambda Bx$$

$$B^{-1}Ax = \lambda x$$

Generalized Eigenvectors

- Another option is to use the power method for generalized eigenvectors:

$$Ax = \lambda Bx$$

$$x \Leftarrow Ax$$

$$x \Leftarrow B \backslash x$$

$$x \Leftarrow x / \|x\|$$

Generalized Eigenvectors

- In our case A is a matrix of rank 1

$$A = (\mu_- - \mu_+)(\mu_- - \mu_+)^{\mathrm{T}} = d\,d^{\mathrm{T}}$$

- And the solution is: $x = B^{-1} d$