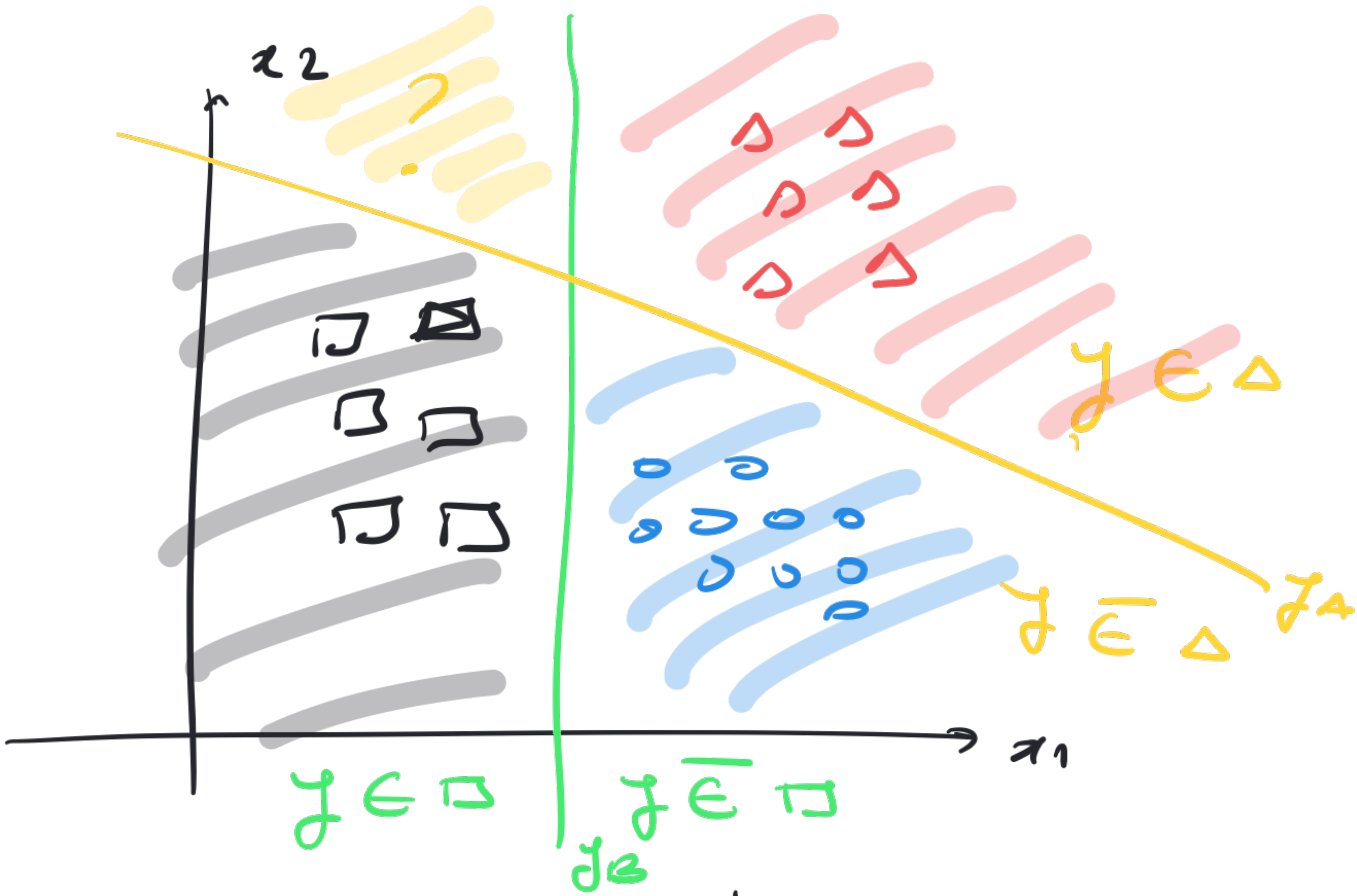
Multiple classes



One-vs-All

(or

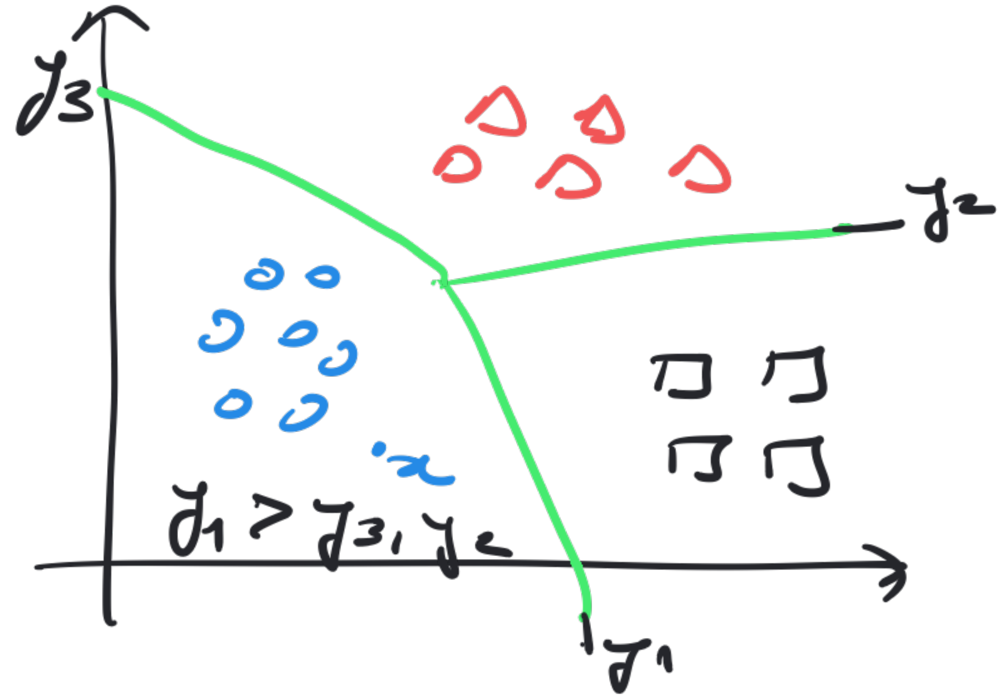
One-vs-Rest)

Single K-class Discriminant :

$$\boxed{y_K(x) = w_K^T \cdot x + b_K}$$

$x \in C_j$ if $y_j(x) > y_K(x) \quad \forall j \neq K$

This approach will create singly convex regions in feature space

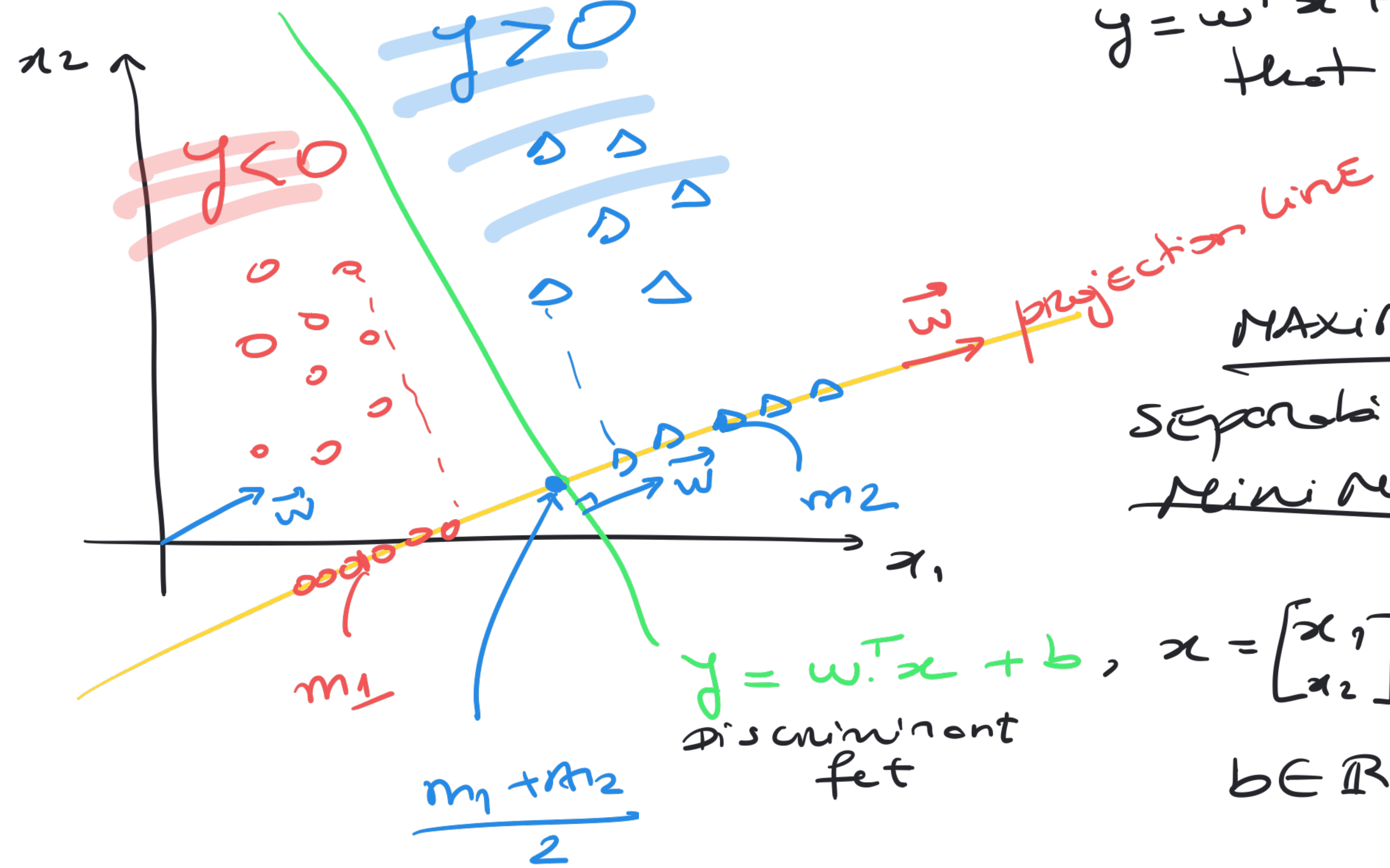


Fisher's Linear Discriminant Analysis

↳ linear classifier, discriminative

↳ supervised linear dimensionality reduction

Find discriminant fct $y = w^T x + b$ such that the projection to orthogonal fct



$y = w^T x + b$, $x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$, $w = \begin{bmatrix} w_1 \\ w_2 \end{bmatrix}$

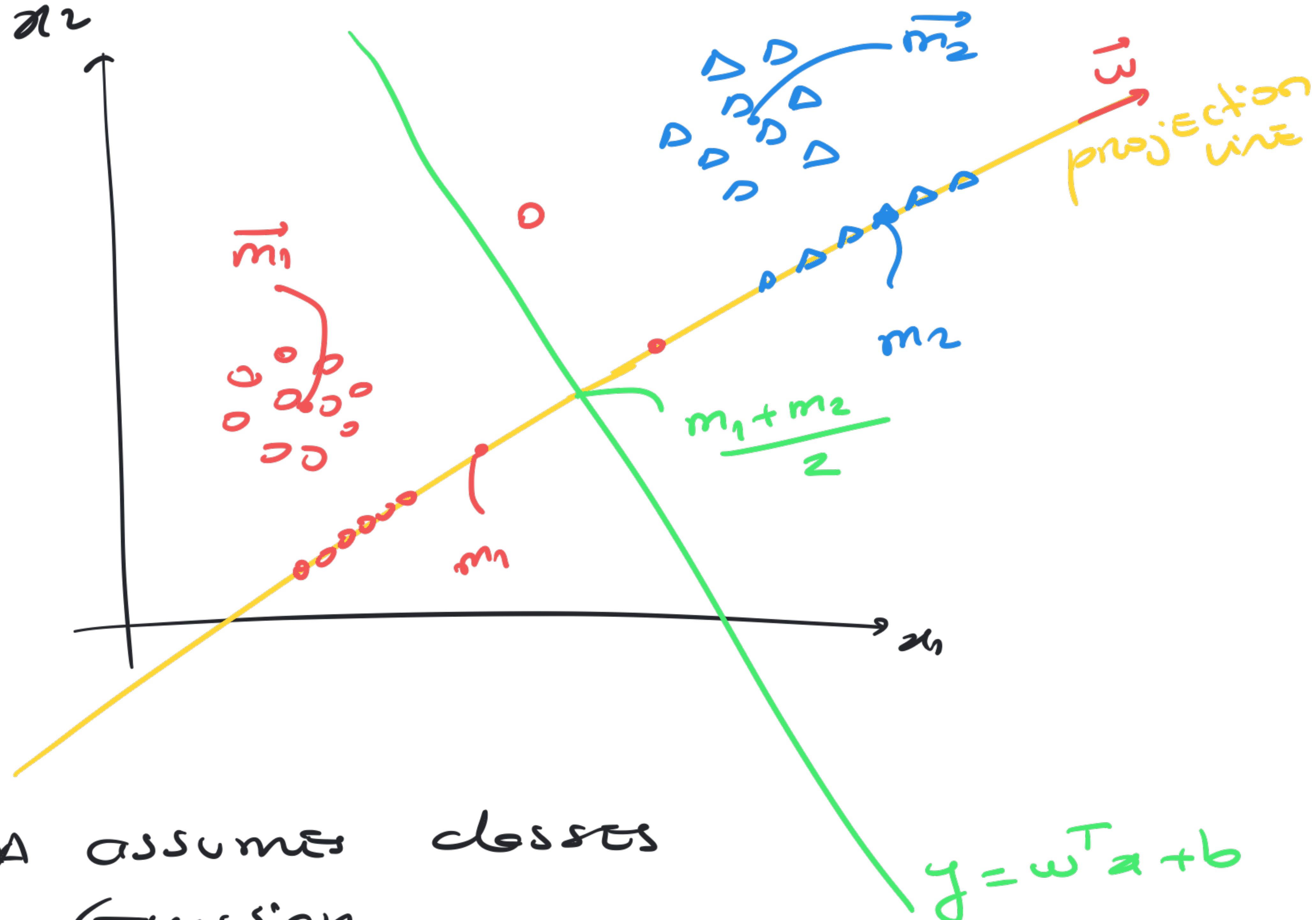
discriminant fct

$b \in \mathbb{R}$

<u>MAXIMIZES</u> class separability and <u>MINIMIZES</u> class compactness

If $y(x) > 0 \implies x \in \triangle$

If $y(x) < 0 \implies x \in \circ$



LDA assumes classes are Gaussian.

OBJECTIVE function

$$J(\vec{\omega}) = \frac{(m_1 - m_2)^2}{s_1^2 + s_2^2}$$

NOTE: this objective fct is assuming we start in a 2-D feature space and project down to 1-D.

$m_i \equiv$ projected mean for class C_i (1-D)

$s_i^2 \equiv$ projected variance for class C_i (1-D)

$$\boxed{\underset{\{\vec{\omega}\}}{\arg\max}\ J(\vec{\omega})}$$

$\vec{m}_i \equiv$ mean for class i in the original space (2-D)

$\Sigma_i \equiv$ covariance for class i in the original space (2-D)

$$m_i = \vec{w}^T \cdot \vec{m_i} \quad \leftarrow \text{vector projection}$$

$$(m_1 - m_2) = \vec{w}^T \cdot (\vec{m_1} - \vec{m_2})$$

For the projected class variance or

within-class variance :

$$\boxed{s_i^2 =} \sum_{n \in C_i} (y(x_n) - m_i)^2$$

$$= \sum_{n \in C_i} (\vec{w}^T \cdot \vec{x}_n + b - \vec{w}^T \cdot \vec{m_i})^2$$

$$= \sum_{n \in C_i} (\vec{w}^T \cdot \vec{x}_n - \vec{w}^T \cdot \vec{m_i})^2$$

$$= \boxed{\vec{w}^T \left(\sum_{n \in C_i} (\vec{x}_n - \vec{m_i})(\vec{x}_n - \vec{m_i})^T \right) \cdot \vec{w}}$$

Substituting,

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

$S_B \equiv$ BETWEEN-class scatter matrix

$$= \frac{\vec{w}^T \underbrace{(\vec{m_1} - \vec{m_2}).(\vec{m_1} - \vec{m_2})^T}_{S_B} . \vec{w}}{\vec{w}^T \underbrace{\left(\sum_{n \in C_1} (\vec{x}_n - \vec{m}_1)(\vec{x}_n - \vec{m}_1)^T + \sum_{n \in C_2} (\vec{x}_n - \vec{m}_2)(\vec{x}_n - \vec{m}_2)^T \right)}_{S_W} . \vec{w}}$$

$S_W \equiv$ within-class scatter MATRIX

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

$$J(\omega) = \frac{\vec{\omega}^T . S_B . \vec{\omega}}{\vec{\omega}^T . S_W . \vec{\omega}} \in \mathbb{R}$$

optimize for $\vec{\omega}$:

$$\frac{\partial J(\vec{\omega})}{\partial \vec{\omega}} = 0 \iff \frac{\cancel{2} . (\vec{\omega}^T . S_W . \vec{\omega}) . S_B . \vec{\omega} - \cancel{2} . (\vec{\omega}^T S_B . \vec{\omega}) . S_W . \vec{\omega}}{(\vec{\omega}^T . S_W . \vec{\omega})^2} = 0$$

$$\iff \frac{\cancel{(\omega^T . S_W . \omega)} \, S_B . \omega}{(\omega^T . S_W . \omega)^{\cancel{2}}} - \frac{(\omega^T S_B . \omega) . S_W . \omega}{(\omega^T . S_W . \omega)^2} = 0$$

$$\iff (\omega^T . S_W . \omega) . S_B . \omega = (\omega^T S_B \, \omega) . S_W . \omega$$

$$\iff S_B . \omega = \boxed{\frac{\omega^T . S_B \, \omega}{\omega^T . S_W . \omega}} . S_W . \omega$$

↳ constant, λ

$$\iff S_B . \omega = \lambda . S_W . \omega$$

$$\iff \boxed{S_W^{-1} . S_B . \omega = \lambda . \omega}$$, if S_W^{-1} exists

↳ CHARACTERISTIC EQUATION

Reminder:

$\vec{V}$ Eigenvector of A

λ is its Eigenvalue

$$A . \vec{V} = \lambda . \vec{V}$$

we pick the eigenvector w with largest eigenvalue of $S_w^{-1} \cdot S_B$

↳ because we want to maximize $J(w)$.

- in practice, if S_w is not invertible (singular matrix), we can diagonally-load it prior to inverting it.

Food for Thought

FEATURE SPACE: 10-D

classes : 2

1 what is the size of $\vec{\omega}$? (YES, 10-D!!)

2 what is the dimensionality of projection subspace?

If we have C classes, we can only project to C-1 dimensions.