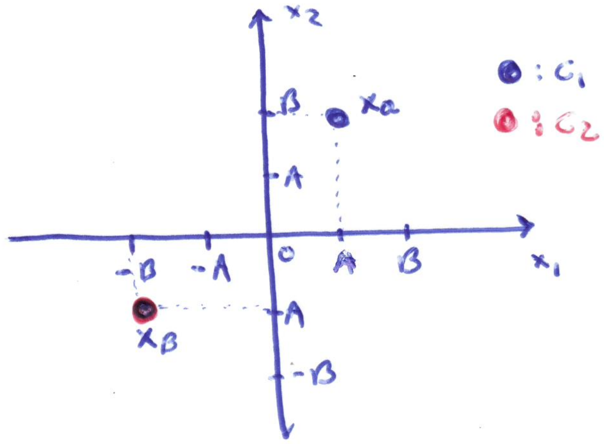
Problem Σ: Consider the following binary classification problem between classes C_1 and C_2 where $\underline{x}_a \in C_1$ and $x_b \in C_2$ such that $\underline{x}_a = [A, B]^T$ and $\underline{x}_B = [-B, -A]^T$ where $0 < A < B$.



Solution: Let $\mathcal{X} = \{\underline{x}_a, \underline{x}_B\}$ be the given dataset where the corresponding class labels are given within the set $\mathcal{Y} = \{y_a, y_B\}$ where $y_a = +1$ and $y_B = -1$.

Formulate the Primal Optimization Problem: (Hard Margin SVM)

$$\boxed{\begin{aligned} &\min_{\underline{w}, b} \ \frac{1}{2} \|\underline{w}\|^2 \\ &\text{s.t.} \quad y_a \cdot (\underline{w}^T \cdot \underline{x}_a + b) \geq 1 \qquad (1) \\ &\qquad\ \ y_B \cdot (\underline{w}^T \cdot \underline{x}_B + b) \geq 1 \end{aligned}}$$

Variables of the convex primal optimization problem are the parameters that define the hard-margin maximizing hyperplane $\boxed{g(\underline{x}) = \underline{w}^T \cdot \underline{x} + b \underset{(2)}{=} 0}$ (2) where $w = [w_1 \ w_2]^T$ and $b \in \mathbb{R}$.

#Variables = Input Dimensionality + 1 = 2 + 1 = 3.

Constraint Functions: The optimization problem defined in Eq. (1) is subject to two linear constraints which may be formulated as:

$$\begin{cases} g_a(\underline{w}, b) \geq 0 \\ g_\beta(\underline{w}, b) \geq 0 \end{cases} \quad (3)$$

where

$$\begin{cases} g_a = y_a(\underline{w}^T \underline{x}_a + b) - 1 \geq 0 \\ \\ g_\beta = y_\beta(\underline{w}^T \underline{x}_\beta + b) - 1 \geq 0 \end{cases} \quad (4)$$

Primal Optimization Problem (Re-formulated):

$$\min_{\underline{w}, b} \ \frac{1}{2} \underline{w}^T \underline{w}$$

$$\text{s.t.} \quad g_a(\underline{w}, b) \geq 0 \qquad (5)$$
$$g_\beta(\underline{w}, b) \geq 0$$

Specific Constraints: Substituting the exact vector representations for the datapoints $\underline{x}_a$ and $\underline{x}_\beta$, the constraint functions g_a, g_β will eventually be formulated as:

* For $\underline{x}_a$: $g_a = y_a\left([w_1 \ w_2]\begin{bmatrix} A \\ B \end{bmatrix} + b\right) - 1 \geq 0$

For $\underline{x}_\beta$: $g_\beta = y_\beta\left([w_1 \ w_2]\begin{bmatrix} -B \\ -A \end{bmatrix} + b\right) - 1 \geq 0$ (6)

* For $\underline{x}_a$:

$$\begin{aligned} g_a &= w_1 A + w_2 B + b - 1 \geq 0 \\ g_\beta &= w_1 B + w_2 A - b - 1 \geq 0 \end{aligned} \quad (7)$$

Hard-Margin SVM Dual Form: The dual optimization problem can be formulated by considering the corresponding Lagrangian function:

$$\boxed{L(\underline{w}, b, \underline{\lambda}) = \frac{1}{2}\underline{w}^T\underline{w} - \sum_{k \in \{\alpha, \beta\}} \lambda_k \circ g_k(\underline{w}, b) \quad (8)}$$

* $\underline{\lambda} = [\lambda_\alpha \; \lambda_\beta]$ is the vector of non-negative ~~matrix~~ Lagrange multipliers.

λ_α : is the Lagrange multiplier associated with constraint g_α.
λ_β : is the Lagrange multiplier associated with constraint g_β.

* The Dual Optimization Problem can be formulated as:

$$\boxed{\begin{aligned} &\max_{\underline{\lambda}} \; \min_{\underline{w}, b} \; L(\underline{w}, b, \underline{\lambda}) \qquad (9) \\ &\text{s.t.} \quad \lambda_k \geq \emptyset, \quad \forall k \in \{\alpha, \beta\} \end{aligned}}$$

Kuhn-Tucker Theorem: The necessary and sufficient conditions for a normal point $(\underline{w}^*, b^*)$ to be an optimum is the existence of $\underline{\lambda}^*$ such that:

KKT CONDITIONS:

(A): Stationarity Conditions

$$\begin{cases} \dfrac{\partial L}{\partial \underline{w}} = \emptyset & (10) \\ \dfrac{\partial L}{\partial b} = \emptyset & (11) \end{cases}$$

(B): Complementarity Slackness

$$\sum_{k \in \{\alpha, \beta\}} \lambda_k \cdot g_k(\underline{w}, b) = \emptyset \quad (12) \text{ or}$$

$$\lambda_k \cdot g_k(\underline{w}, b) = \emptyset, \ \forall k \in \{\alpha, \beta\} \quad (13)$$

(Γ): Primal Feasibility:

$$g_k(\underline{w}, b) \geq \emptyset, \ \forall k \in \{\alpha, \beta\} \quad (14)$$

(Δ): Dual Feasibility:

$$\lambda_k \geq 0, \ \forall k \in \{\alpha, \beta\} \quad (15)$$

★ The kth linear inequality constraint may be expressed as:

$$\boxed{g_k(\underline{w}, b) = y_k \cdot (\underline{w}^T \underline{x}_k + b) - 1, \ \forall k \in \{\alpha, \beta\}} \quad (16)$$

▷ We need to form its partial derivatives with respect to both $\underline{w}$ and b as:

(i): $$\frac{\partial g_k}{\partial \underline{w}} = y_k \cdot \frac{\partial}{\partial \underline{w}} \{\underline{w}^T \underline{x}_k + b\} = y_k \cdot \frac{\partial}{\partial \underline{w}} \{\underline{w}^T \underline{x}_k\} = y_k \underline{x}_k, \ \forall k \in \{\alpha, \beta\} \quad (17)$$

(ii): $$\frac{\partial g_k}{\partial b} = y_k \cdot \frac{\partial}{\partial b} \{\underline{w}^T \underline{x}_k + b\} = y_k \cdot \frac{\partial}{\partial b} \{b\} = y_k, \ \forall k \in \{\alpha, \beta\} \quad (18)$$

KKT Conditions Exploration:

(A): Stationarity Conditions:

Eq. (10) yields that: $\frac{\partial L}{\partial \underline{w}} = \mathbf{0} \Rightarrow \frac{\partial}{\partial \underline{w}} \left\{ \frac{1}{2} \underline{w}^T \underline{w} - \sum_{k \in \{\alpha, \beta\}} \lambda_k \, g_k(\underline{w}, b) \right\} = \mathbf{0} \rightarrow$

$$\frac{1}{2} \frac{\partial}{\partial \underline{w}} \{ \underline{w}^T \underline{w} \} - \sum_{k \in \{\alpha, \beta\}} \lambda_k \frac{\partial}{\partial \underline{w}} g_k(\underline{w}, b) = \mathbf{0} \Rightarrow$$ (Having in mind Eq. (17))

$$\frac{1}{\cancel{2}} \cdot \cancel{2} \cdot \underline{w} - \sum_{k \in \{\alpha, \beta\}} \lambda_k y_k \underline{x}_k = \mathbf{0} \rightarrow \boxed{\underline{w}^* = \sum_{k \in \{\alpha, \beta\}} \lambda_k^* y_k \underline{x}_k \quad (19)}$$

Eq. (19) may also be written as: $\underline{w}^* = \lambda_\alpha y_\alpha \underline{x}_\alpha + \lambda_\beta y_\beta \underline{x}_\beta \rightarrow$

$$\boxed{\underline{w}^* = \lambda_\alpha^* \underline{x}_\alpha - \lambda_\beta^* \underline{x}_\beta \quad (20)}$$

Eq. (11) yields that: $\frac{\partial L}{\partial b} = 0 \Rightarrow - \sum_{k \in \{\alpha, \beta\}} \lambda_k^* \frac{\partial}{\partial b} g_k(\underline{w}, b) = \mathbf{0} \rightarrow$ (Eq. 18)

$$\boxed{\sum_{k \in \{\alpha, \beta\}} \lambda_k^* y_k = \mathbf{0} \quad (21)}$$

Eq. (21) may be reformulated as: $\lambda_\alpha y_\alpha + \lambda_\beta y_\beta = \mathbf{0} \rightarrow$

$$\boxed{\lambda_\alpha^* - \lambda_\beta^* = \mathbf{0} \quad (22)}$$

From Eq. (22) we may deduce that: $\boxed{\lambda^* = \lambda_\alpha^* = \lambda_\beta^* \quad (23)}$

Taking into consideration Eqs. (20) and (23), we may conclude that:

$$\boxed{\underline{w}^* = \lambda^* (\underline{x}_\alpha - \underline{x}_\beta) \quad (24)}$$

(Β): Complementarity Slackness:

$$\boxed{\begin{aligned} \lambda^{*} \cdot g_{a}(\underline{w}, b) &= \emptyset \quad (25) \\ \lambda^{*} \cdot g_{\beta}(\underline{w}, b) &= \emptyset \quad (26) \end{aligned}}$$

(Γ): Primal Feasibility:

$$\boxed{\begin{aligned} g_{a}(\underline{w}, b) &\geqslant \emptyset \quad (27) \\ g_{\beta}(\underline{w}, b) &\geqslant \emptyset \quad (28) \end{aligned}}$$

(Δ): Dual Feasibility:

$$\boxed{\lambda^{*} \geqslant \emptyset \quad (29)}$$

$\boxed{\rightarrow}$

Taking into consideration the Complementarity Slackness Karush-Kuhn-Tucker conditions, we have that:

(i): For active constraints $\boxed{(\lambda^{*} = 0)}$, we may deduce that:

$$\boxed{\begin{cases} g_{a} > \emptyset \\ g_{\beta} > \emptyset \end{cases} \quad (30)} \rightarrow \begin{cases} y_{a}(\underline{w}^{T*} \underline{x}_{a} + b^{*}) - 1 \geqslant 0 \\ y_{\beta}(\underline{w}^{T*} \underline{x}_{\beta} + b^{*}) - 1 \geqslant 0 \end{cases} \rightarrow$$

$$\begin{cases} y_{a}(\underline{w}^{T*} \underline{x}_{a} + b^{*}) \geqslant 1 \\ y_{\beta}(\underline{w}^{T*} \underline{x}_{\beta} + b^{*}) \geqslant 1 \end{cases} \rightarrow \begin{cases} \underline{w}^{T*} \underline{x}_{a} + b^{*} \geqslant 1 \\ -(\underline{w}^{T*} \underline{x}_{\beta} + b^{*}) \geqslant 1 \end{cases} \Rightarrow$$

$$\boxed{\begin{cases} \underline{w}^{T*} \underline{x}_{a} + b^{*} \geqslant 1 \\ \underline{w}^{T*} \underline{x}_{\beta} + b^{*} \leqslant -1 \end{cases} \quad (31)}$$

According to Eq.(24), when $\lambda^{*} = \emptyset$, $\underline{w}^{*} = \emptyset$. Thus, Eqs.(31), yield that:

$$\begin{cases} b^{*} \geqslant 1 \\ b^{*} \leqslant -1 \end{cases} \Rightarrow \text{IMPOSSIBLE !!!}$$

(ii): For inactive constraints $(\lambda^* > \emptyset)$, we may deduce that:

$$\boxed{\begin{cases} g_a = \emptyset \\ g_B = \emptyset \end{cases} \quad (32)} \Rightarrow \begin{cases} y_a \cdot (\underline{w}^{T*} \underline{x}_a + b^*) - 1 = \emptyset \\ y_B \cdot (\underline{w}^{T*} \underline{x}_B + b^*) - 1 = \emptyset \end{cases} \longrightarrow$$

$$\begin{cases} y_a (\underline{w}^{*T} \underline{x}_a + b^*) = 1 \\ y_B (\underline{w}^{*T} \underline{x}_B + b^*) = 1 \end{cases} \rightarrow \boxed{\begin{cases} \underline{w}^{T*} \underline{x}_A + b^* = 1 \quad (33_A) \\ \underline{w}^{T*} \cdot \underline{x}_B + b^* = -1 \quad (33_B) \end{cases}}$$

⊛ Performing pairwise subtraction between Eqs. (33_A) and (33_B), we

$$\underline{w}^{*T} \cdot \underline{x}_a - \underline{w}^{*T} \cdot \underline{x}_B = 2 \rightarrow$$

$$\boxed{\underline{w}^{*T} \cdot (\underline{x}_a - \underline{x}_B) = 2 \quad (34)}$$

⊛ Taking into account Eq. (24), Eq. (34) yields that:

$$\lambda^* (\underline{x}_a - \underline{x}_B)^T (\underline{x}_a - \underline{x}_B) = 2 \Rightarrow$$

$$\lambda^* \| \underline{x}_a - \underline{x}_B \|^2 = 2 \Rightarrow \boxed{\lambda^* = \frac{2}{\| \underline{x}_a - \underline{x}_B \|^2} \quad (35)}$$

⊛ Having in mind that: $\underline{x}_a = [A, B]^T$ and $\underline{x}_B = [-B, -A]^T$, we may write that:

$$\underline{x}_a - \underline{x}_B = \begin{bmatrix} A \\ B \end{bmatrix} - \begin{bmatrix} -B \\ -A \end{bmatrix} = \begin{bmatrix} A+B \\ A+B \end{bmatrix} = [A+B \quad A+B]^T \quad (36)$$

⊛ Therefore, Eq. (35), gives that:

$$\lambda^* = \frac{2}{(A+B)^2 + (A+B)^2} = \frac{2}{2(A+B)^2} \Rightarrow \boxed{\lambda^* = \frac{1}{(A+B)^2} \quad (37)}$$

⊛ Thus, both data points $\underline{x}_a$ and $\underline{x}_B$ are support vectors.

⊛ Performing pairwise addition between Eqs. (33_A) and (33_B), we get that:

$$\underline{w}^{*T} \circ (\underline{x}_a + \underline{x}_\beta) + 2b^* = \emptyset \Rightarrow$$

$$2b^* = -\underline{w}^{*T}(\underline{x}_a + \underline{x}_\beta) \rightarrow$$

$$\boxed{b^* = -\frac{1}{2} \circ \underline{w}^{*T} \circ (\underline{x}_a + \underline{x}_\beta) \quad (38)}$$

⊛ We have already established that $\underline{w}^* = \lambda^*(x_a - x_\beta) \Rightarrow$

$$\boxed{\underline{w}^* = \frac{2}{\|\underline{x}_a - \underline{x}_\beta\|^2} \circ (\underline{x}_a - \underline{x}_\beta) \quad (39)}$$

⊛ Substituting Eq. (39) into Eq. (38), we get that:

$$b^* = -\frac{1}{\cancel{2}} \circ \cancel{2} \circ \frac{1}{\|\underline{x}_a - \underline{x}_\beta\|^2} \circ (\underline{x}_a - \underline{x}_\beta)^T (\underline{x}_a + \underline{x}_\beta) \rightarrow$$

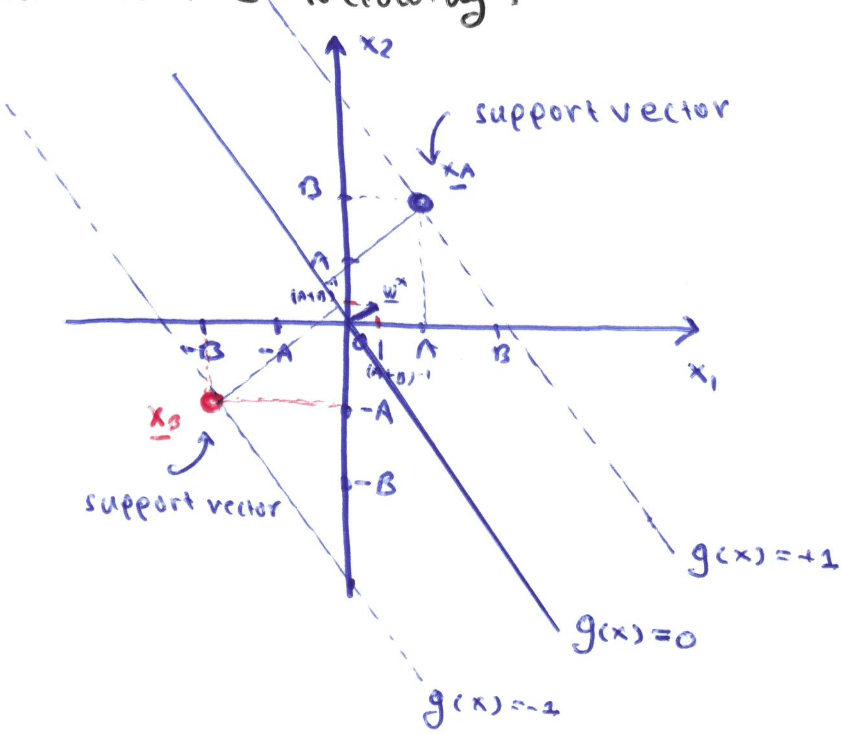
$$b^* = -\frac{1}{\|\underline{x}_a - \underline{x}_\beta\|^2} \circ \left(\|\underline{x}_a\|^2 - \|\underline{x}_\beta\|^2 \right) \xrightarrow[\|\underline{x}_\beta\|^2 = A^2 + B^2]{\|\underline{x}_a\|^2 = A^2 + B^2}$$

$$\boxed{b^* = \emptyset \quad (40)}$$

⊛ Eq. (39) suggests: $\underline{w}^* = \frac{\cancel{2}}{\cancel{2}(A+B)^2} \circ \begin{bmatrix} A+B \\ A+B \end{bmatrix} \Rightarrow$

$$\boxed{\underline{w}^* = \begin{bmatrix} \frac{1}{A+B} & \frac{1}{A+B} \end{bmatrix} \quad (41)}$$

(4) The schematic representation of the hard-margin maximizing hyperplane is the following:



⊛ Eq. (1a) states that : $\boxed{\underline{w} = \sum_{k \in \{\alpha,\beta\}} \lambda_k y_k \underline{x}_k}$

⊛ Eq. (21) states that : $\boxed{\sum_{k \in \{\alpha,\beta\}} \lambda_k y_k = 0}$

▶ Forming the Dual Optimization Problem requires substituting the expression for $\underline{w}$ ($\underline{w}^*$) into the Lagrangian expression (Eq. 8):

$$L(\underline{w}, b, \underline{\lambda}) = \underbrace{\frac{1}{2}\underline{w}^T\underline{w}}_{Q_1} - \underbrace{\sum_{k \in \{\alpha,\beta\}} \lambda_k \cdot g_k(\underline{w}, b)}_{Q_2}$$

▶ Let $Q_2 = \sum_{k \in \{\alpha,\beta\}} \lambda_k g_k(\underline{w}, b)$ where $g_k(\underline{w}, b) = y_k(\underline{w}^T\underline{x}_k + b) - 1$,

which by substituting Eq. (1a), yields:

$$g_k = y_k \left\{ \left[\sum_{r \in \{\alpha,\beta\}} \lambda_r y_r \underline{x}_r^T \right] \underline{x}_k + b \right\} - 1 \Rightarrow$$

$$g_k = y_k \left\{ \sum_{r \in \{\alpha,\beta\}} \lambda_r y_r \underline{x}_r^T \underline{x}_k + b \right\} - 1 \Rightarrow$$

$$\boxed{g_k = \sum_{r \in \{\alpha,\beta\}} \lambda_r y_k y_r x_k^T x_r + b y_k - 1 \quad (42)}$$

▶ Thus, by substituting Eq. (42) into Q_2, we get:

$$Q_2 = \sum_{k \in \{\alpha,\beta\}} \lambda_k \left\{ \sum_{r \in \{\alpha,\beta\}} \lambda_r y_k y_r x_k^T x_r + b y_k - 1 \right\} \Rightarrow$$

$$\boxed{Q_2 = \sum_{k \in \{\alpha,\beta\}} \sum_{r \in \{\alpha,\beta\}} \lambda_k \lambda_r y_k y_r x_k^T x_r + \sum_{k \in \{\alpha,\beta\}} b \lambda_k y_k - \sum_{k \in \{\alpha,\beta\}} \lambda_k \quad (43)}$$

▷ Substituting Eq. (19) into Q_1, we get:

$$Q_1 = \frac{1}{2}\left(\sum_{k\in\{a,b\}} \lambda_k y_k \underline{x}_k^T\right)\left(\sum_{r\in\{a,b\}} \lambda_r y_r \underline{x}_r\right) \Rightarrow$$

$$\boxed{Q_1 = \frac{1}{2}\sum_{k\in\{a,b\}}\sum_{r\in\{a,b\}} \lambda_k \lambda_r y_k y_r \underline{x}_k^T \underline{x}_r \quad (44)}$$

▷ Taking into consideration Eq. (21), we may re-write Q_2 as

$$\boxed{Q_2 = \sum_{k\in\{a,b\}}\sum_{r\in\{a,b\}} \lambda_k \lambda_r y_k y_r \underline{x}_k^T \underline{x}_r - \sum_{k\in\{a,b\}} \lambda_k \quad (45)}$$

⊛ In this context, $L = Q_1 - Q_2 \Rightarrow$

$$\boxed{L(\underline{\lambda}) = \sum_{k\in\{a,b\}} \lambda_k - \frac{1}{2}\sum_{k\in\{a,b\}}\sum_{r\in\{a,b\}} \lambda_k \lambda_r y_k y_r \underline{x}_k^T \underline{x}_r \quad (46)}$$

▷ Dual Optimization Problem:

$$\max_{\underline{\lambda}} L(\underline{\lambda}) = \sum_{k\in\{a,b\}} \lambda_k - \frac{1}{2}\sum_{k\in\{a,b\}}\sum_{r\in\{a,b\}} \lambda_k \lambda_r y_k y_r \underline{x}_k^T \underline{x}_r$$

$$\text{s.t.} \quad \sum_{k\in\{a,b\}} \lambda_k y_k = 0 \qquad (47)$$

⊛ The new form of the Lagrangian function will be given as:

$$L(\lambda) = \lambda_a + \lambda_b - \frac{1}{2}\left\{\lambda_a^2 \underline{x}_a^T \underline{x}_a - \lambda_a \lambda_b \underline{x}_a^T \underline{x}_b - \lambda_a \lambda_b \underline{x}_b^T \underline{x}_a + \lambda_b^2 \underline{x}_b^T \underline{x}_b\right\} \rightarrow$$

$$\boxed{L(\lambda) = \lambda_a + \lambda_b - \frac{1}{2}\left\{\lambda_a^2 \underline{x}_a^T \underline{x}_a - 2\lambda_a \lambda_b \underline{x}_a^T \underline{x}_b + \lambda_b^2 \underline{x}_b^T \underline{x}_b\right\} \quad (48)}$$

- Taking into consideration the condition denoted by Eq. (21),

$$\sum_{k \in \{a, B\}} \lambda_k y_k = 0 \Rightarrow \lambda_a - \lambda_B = 0 \Rightarrow \boxed{\lambda_a = \lambda_B},$$

we may write Eq. (48) as:

$$L(\lambda) = 2\lambda - \frac{1}{2}\left\{ \lambda^2 \underline{x_a}^T \underline{x_a} - 2\lambda^2 \underline{x_a}^T \underline{x_B} + \lambda^2 \underline{x_B}^T \underline{x_B} \right\} \Rightarrow$$

$$L(\lambda) = 2\lambda - \frac{1}{2}\lambda^2 \left\{ \underline{x_a}^T \underline{x_a} - 2\underline{x_a}^T \underline{x_B} + \underline{x_B}^T \underline{x_B} \right\} \rightarrow$$

$$\boxed{L(\lambda) = 2\lambda - \frac{1}{2}\lambda^2 \| \underline{x_a} - \underline{x_B} \|^2 \quad (49)}$$

- Therefore, the Dual Optimization Problem reduces to:

$$\boxed{\max_{\lambda \geq 0} L(\lambda) = 2\lambda - \frac{1}{2}\lambda^2 \| \underline{x_a} - \underline{x_B} \|^2 \quad (50)}$$

- Imposing First Order Conditions on the updated Lagrangian Function, yields:

$$\frac{dL}{d\lambda} = 0 \Rightarrow$$

$$\frac{d}{d\lambda}\left\{ 2\lambda - \frac{1}{2}\lambda^2 \| \underline{x_a} - \underline{x_B} \|^2 \right\} = 0 \Rightarrow$$

$$2 - \lambda \| \underline{x_a} - \underline{x_B} \|^2 = 0 \Rightarrow$$

$$\boxed{\lambda^* = \frac{2}{\| \underline{x_a} - \underline{x_B} \|^2} \quad (51)}$$

- We have already established that the determination of λ^* can ultimately determine the rest of the parameters of the problem.