Support Vector Machines

Prepared by Joseph Bakarji

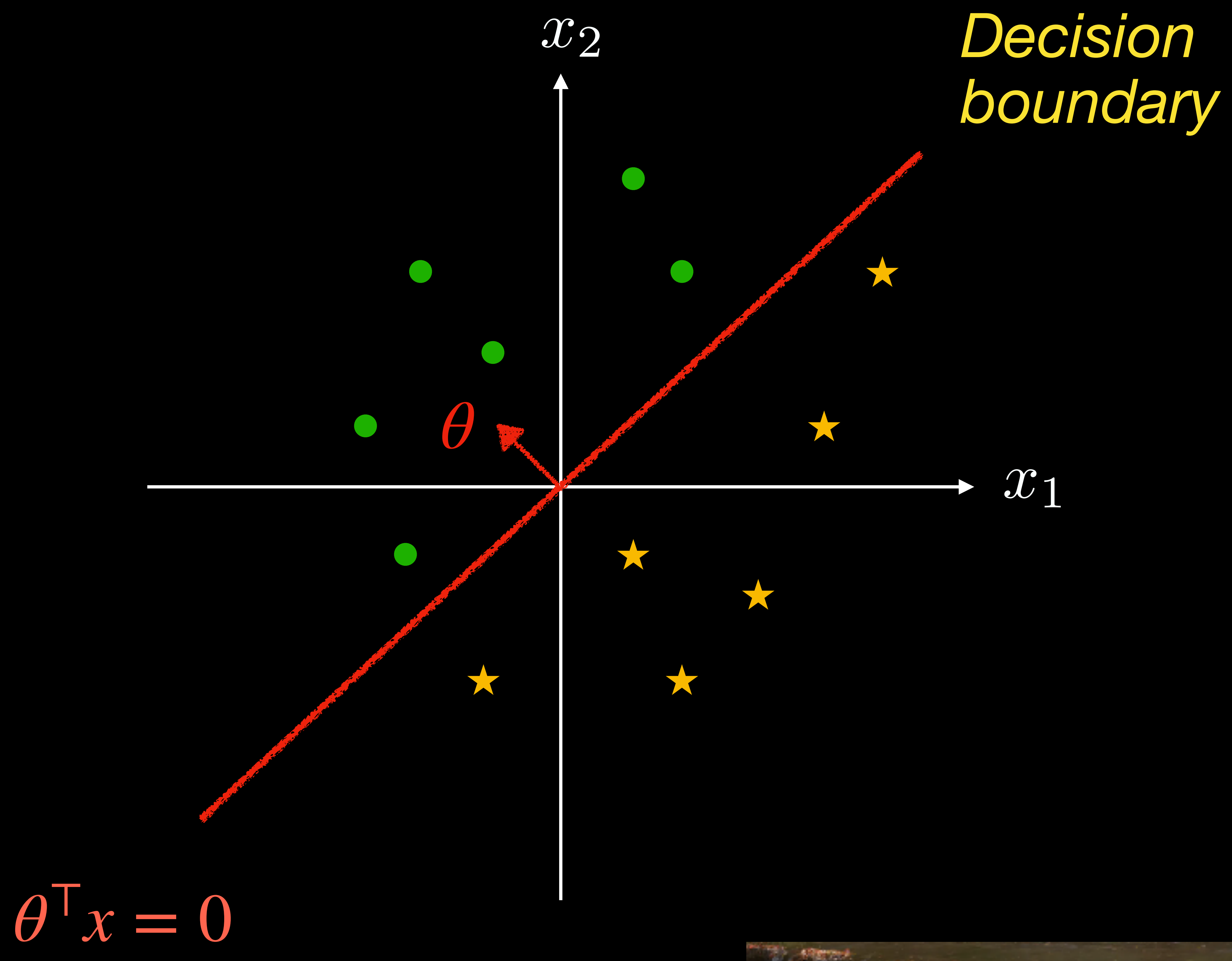
Classification

$x = [x_1, x_2]$

x_1	x_2	y
-2	-1	●
3	1	★
2	3	●
1	-1	★
	⋮	

★ 1

● 0



how confident?

score

$\theta^\top x$

how correct?

margin

$(\theta^\top x)y$

For $y \in [1, -1]$

Logistic Regression

$h_\theta(x) = \sigma(\theta^\top x)$

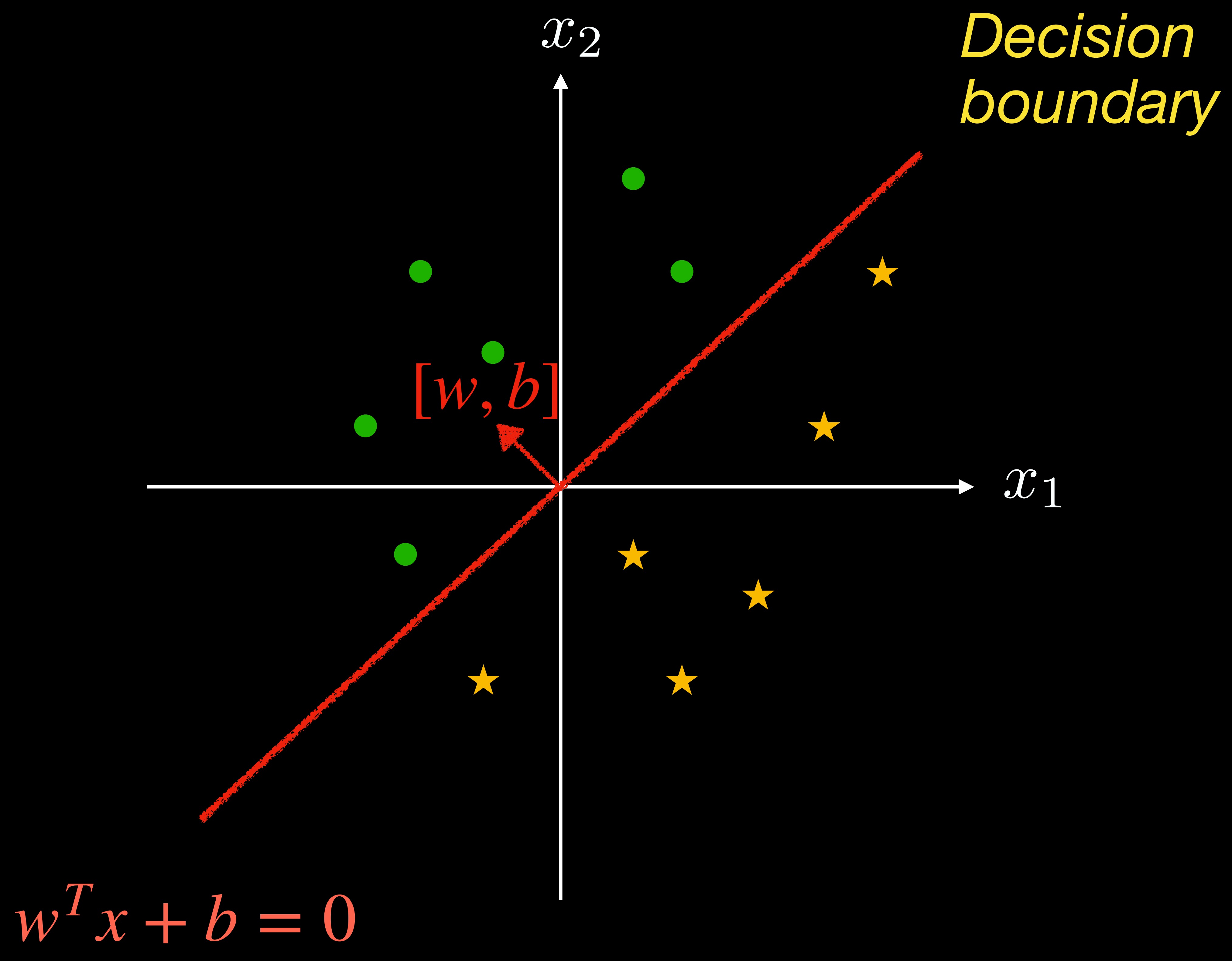
Classification

$x = [x_1, x_2]$

x_1	x_2	y	
-2	-1	●	1
3	1	★	-1
2	3	●	1
1	-1	★	-1
	⋮		

★ -1

● 1



how confident?

score

$w^T x + b$

how correct?

margin

$(w^T x + b) y$

For $y \in [1, -1]$

Classifier

$h_{w,b}(x) = g(w^T x + b)$

What is g?

Functional Margin

Confidence and Correctness

Functional margin

$$\hat{\gamma}^{(i)} = (w^T x^{(i)} + b)\, y^{(i)}$$

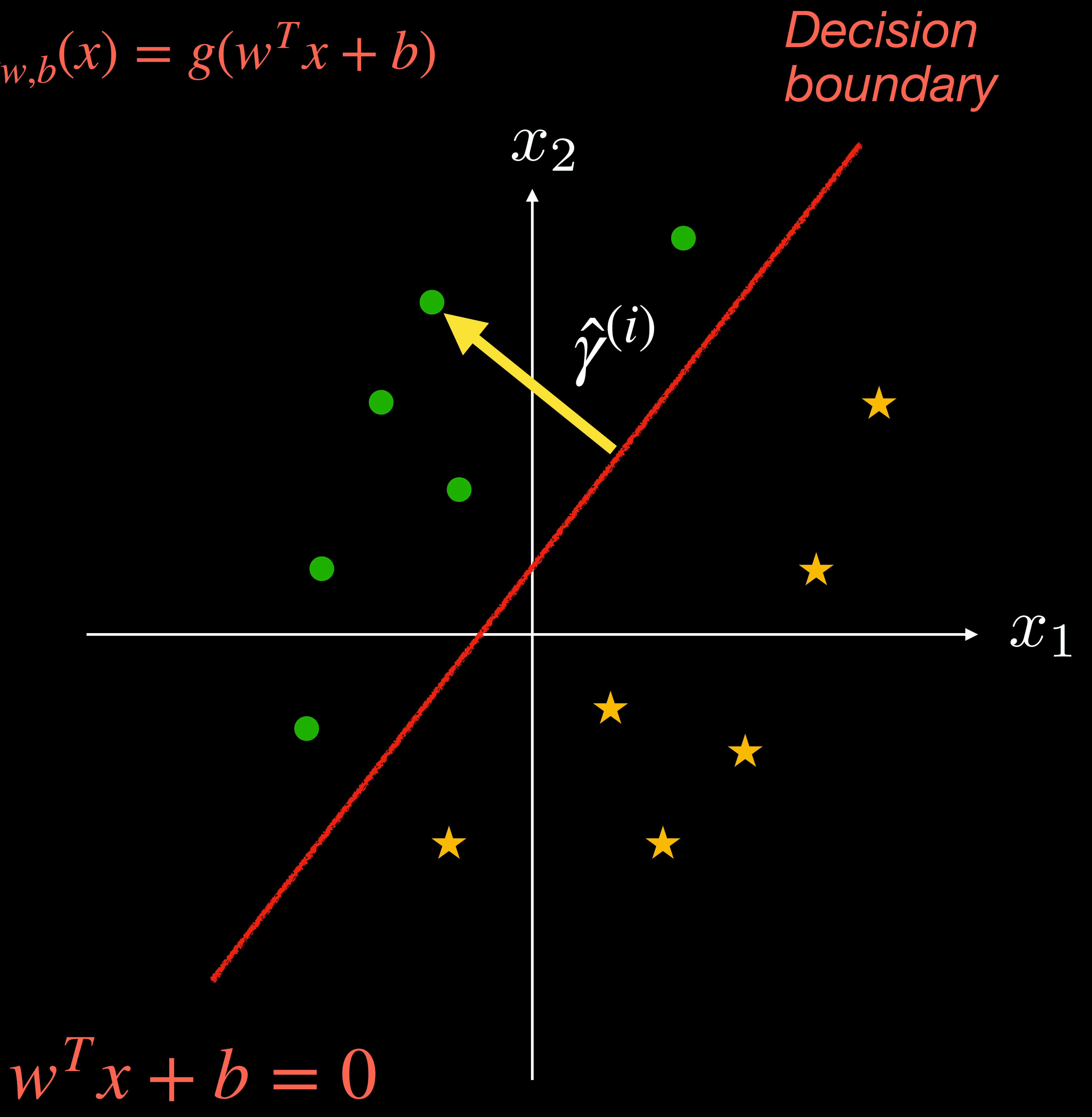
For $y \in [1, -1]$

Functional margin
with respect to $S = \{(x^{(i)}, y^{(i)}\}_{i=1}^{n}$

$$\hat{\gamma} = \min_{i=1,\ldots,n} \hat{\gamma}^{(i)}$$

Problem with Classifier?

$h_{w,b}(x) = g(w^T x + b)$



Geometric Margin

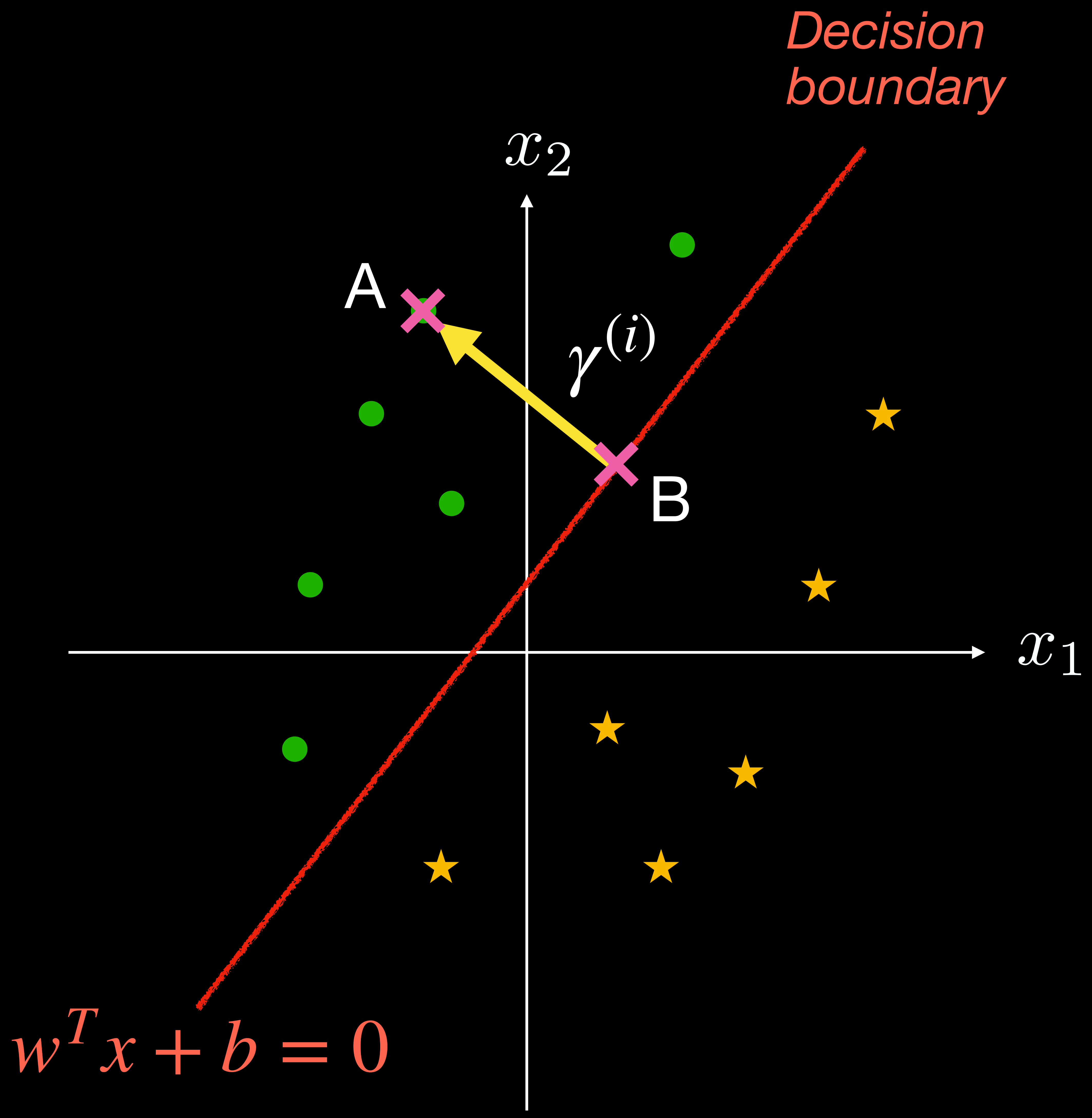
What's the Value of
$\gamma^{(i)}$ (distance to Decision Boundary)?

$\frac{w}{||w||}$ is the a unit length vector

Point "A" represents $x^{(i)}$

What's "B"?

$$x^{(i)} - \gamma^{(i)} \cdot \frac{w}{||w||}$$



Lies on the boundary -> Satisfies equation of Decision Boundary

Support Vector Machines

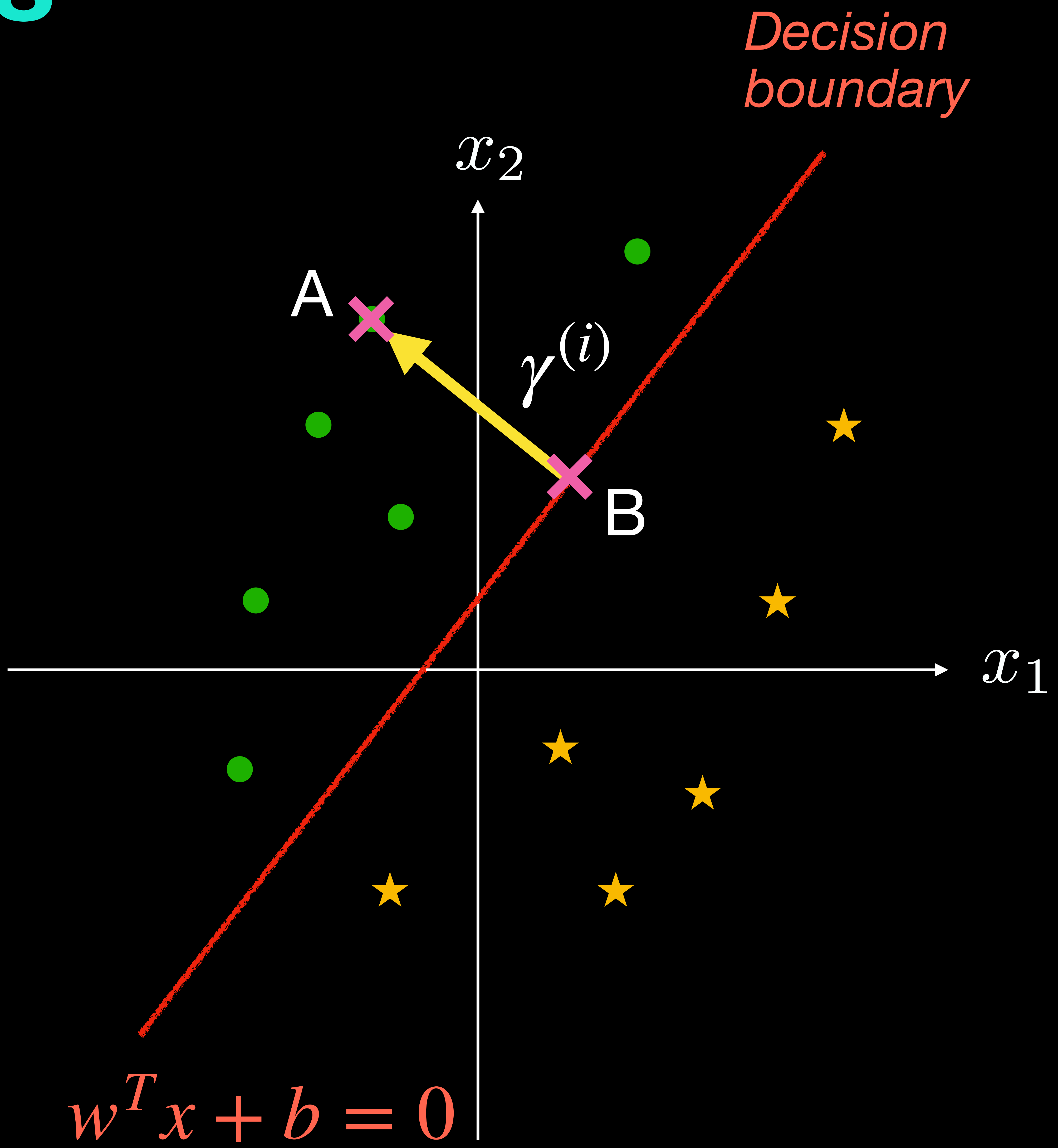
$$w^T\left(x^{(i)} - \gamma^{(i)} \cdot \frac{w}{\|w\|}\right) + b = 0$$

Solve for $\gamma^{(i)}$

$$\gamma^{(i)} = \left(\frac{w}{\|w\|}\right)^T x^{(i)} + \frac{b}{\|w\|}$$

More generally (for non-positive $y^{(i)}$)

$$\gamma^{(i)} = y^{(i)}\left(\left(\frac{w}{\|w\|}\right)^T x^{(i)} + \frac{b}{\|w\|}\right)$$



When $\|w\| = 1$, Geom. = Func. Margin

Support Vector Machines

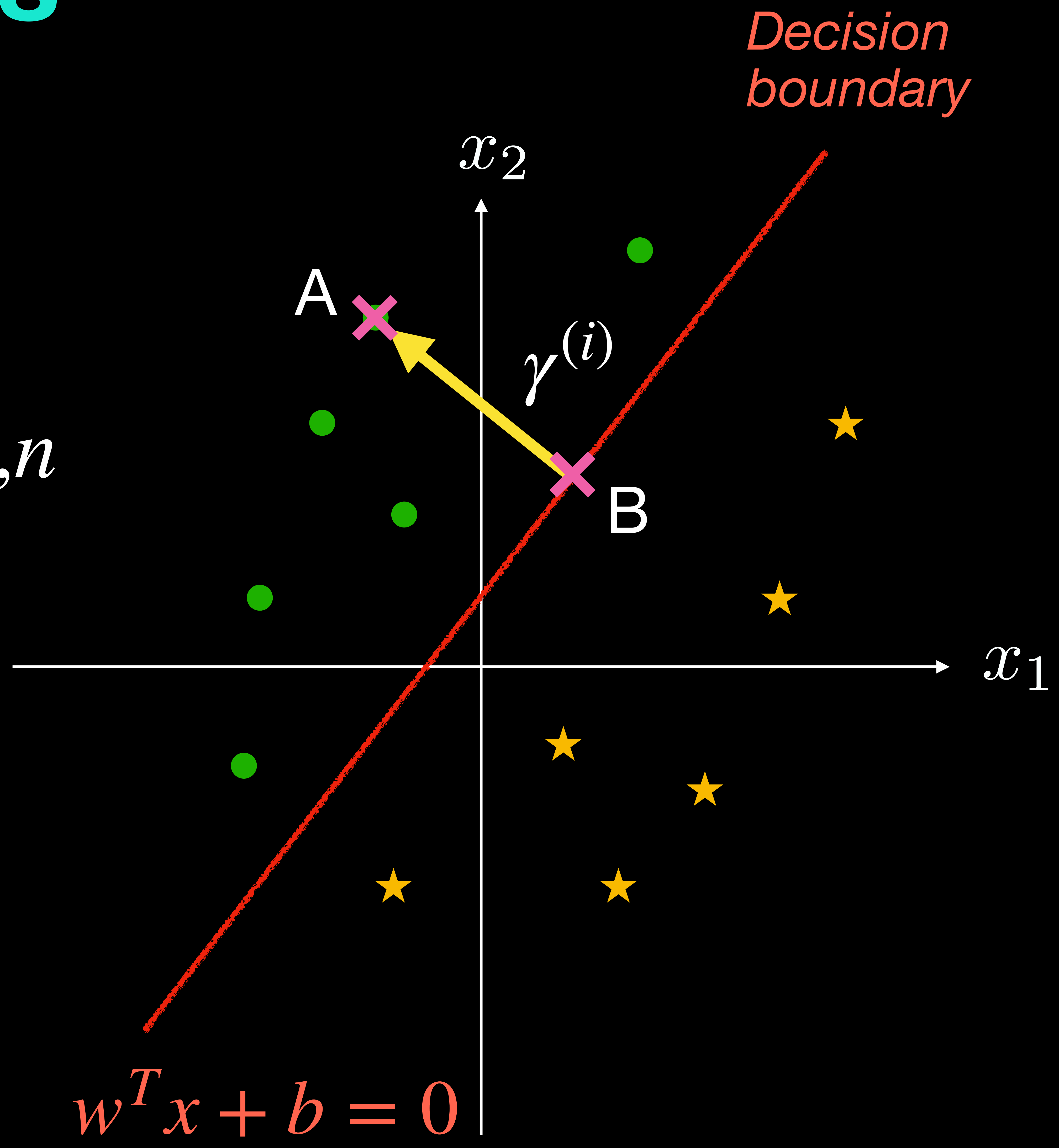
$$\max_{\gamma,w,b} \gamma$$

s.t $\quad y^{(i)}(w^T x^{(i)} + b) \geq \gamma, \;\; for \;\; i = 1,..,n$

$$\|w\| = 1$$

....

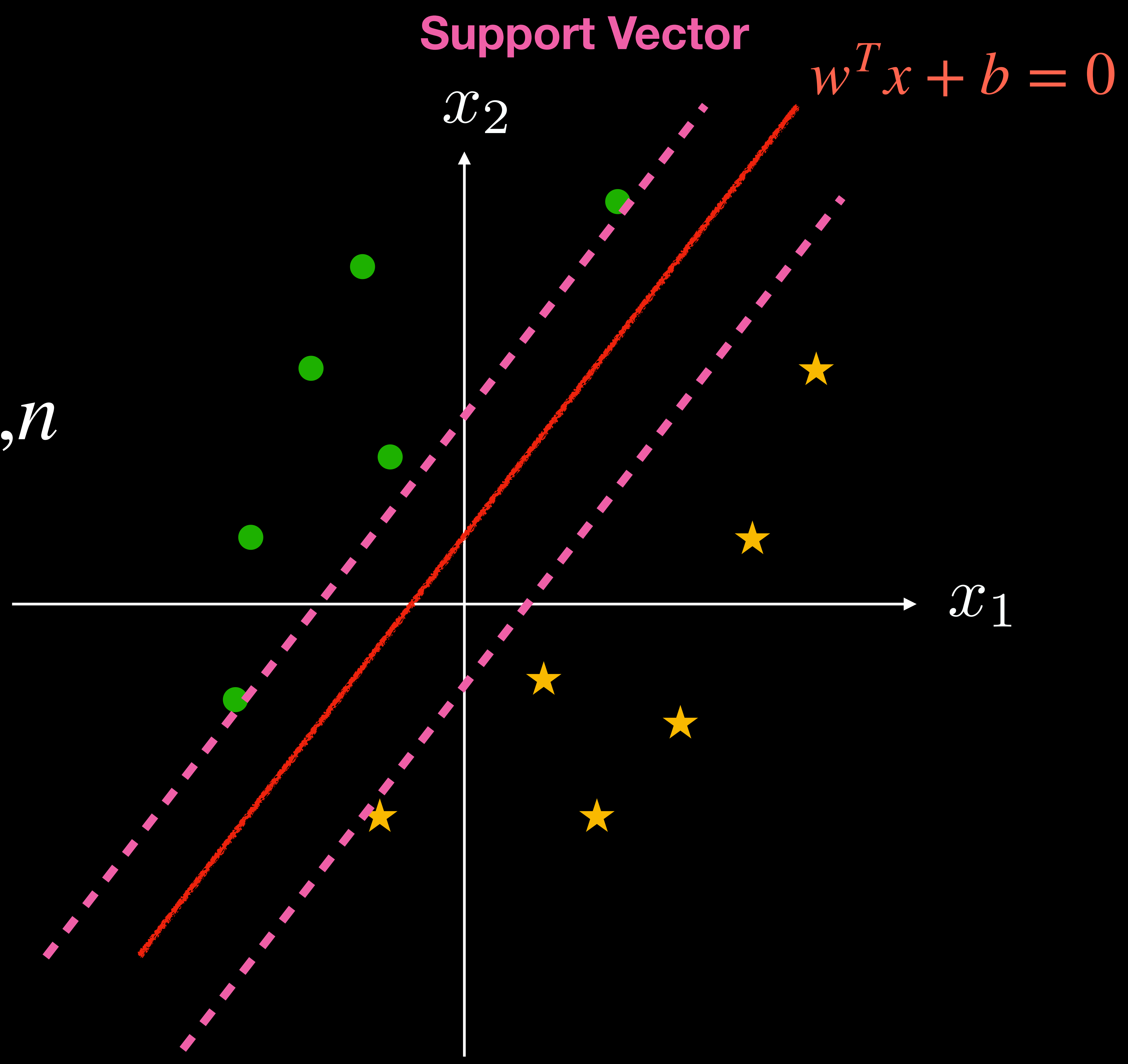
Find a 'nicer' optimization problem using Lagrange Multipliers



Support Vector Machines

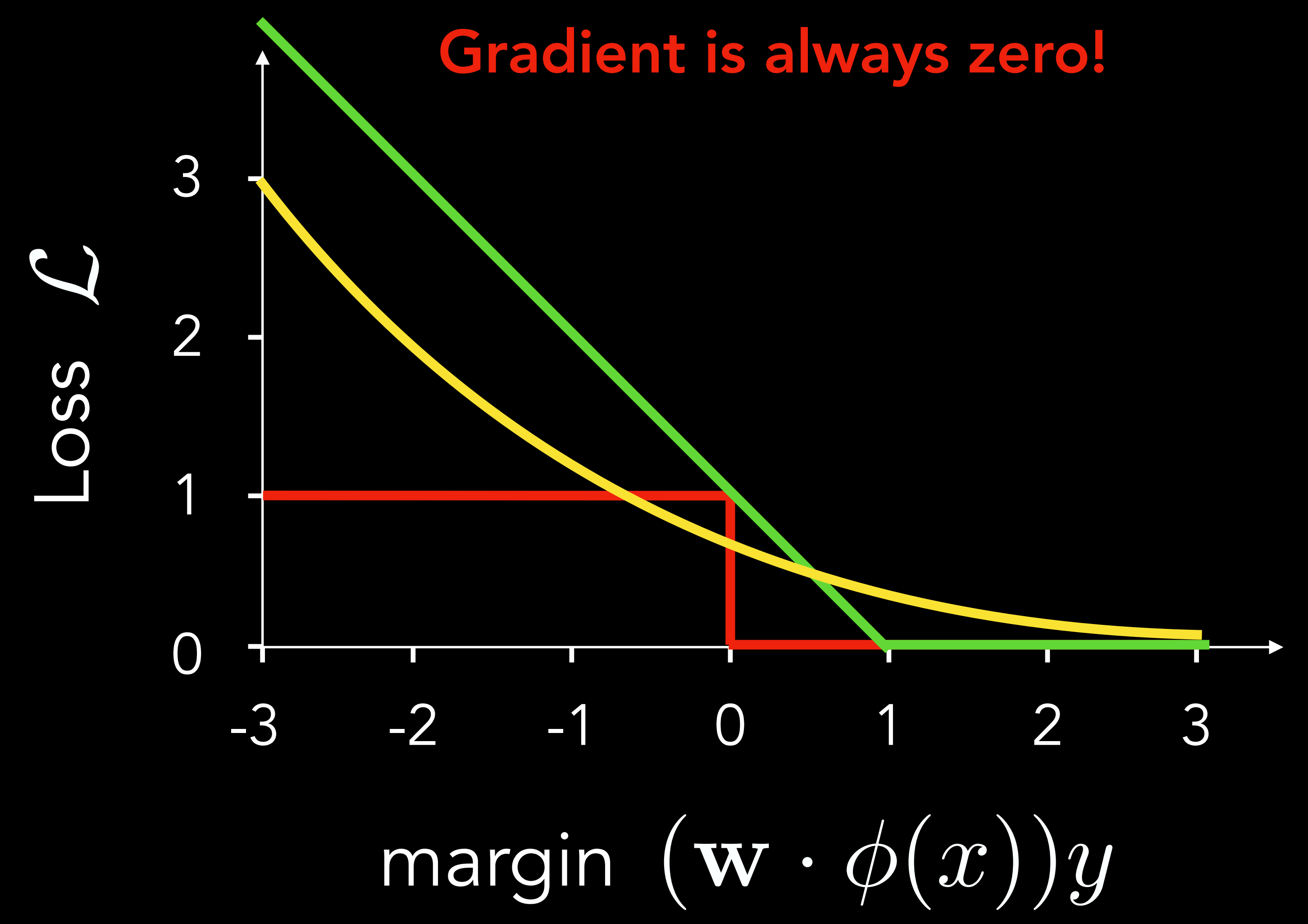
$$\max_{\gamma,w,b} \frac{1}{2}\|w\|^2$$

s.t $\quad y^{(i)}(w^T x^{(i)} + b) \geq 1, \;\; for \;\; i = 1,..,n$



This problem can be expressed in terms of $\langle x^{(i)}, x^{(j)} \rangle$

Classification losses



$$\mathcal{L}_{\text{0-1}} = \mathbf{1}\left[(\mathbf{w} \cdot \phi(x))\, y \leq 0\right]$$

$$\mathcal{L}_{\text{hinge}} = \max\left\{1 - (\mathbf{w} \cdot \phi(x))\, y,\ 0\right\}$$

Support Vector Machines

$$\mathcal{L}_{\text{logistic}} = \log\left(1 + e^{-(\mathbf{w} \cdot \phi(x))y}\right)$$

Logistic Regression

Example MCQ

- What is the main purpose of logistic regression in machine learning?
  - A. To predict continuous values
  - B. To classify data into multiple categories
  - C. To calculate probabilities for binary classification tasks
  - D. To cluster data points

Example MCQ

- Which of the following is an advantage of using logistic regression over SVM for classification?
  - A. Logistic regression can handle non-linear boundaries easily.
  - B. Logistic regression provides probabilities for class predictions.
  - C. Logistic regression is computationally more complex than SVM.
  - D. Logistic regression is only suitable for regression problems.

Example MCQ

- What role does the “kernel trick” play in Support Vector Machines?
  - A. It transforms the data into a higher-dimensional space to handle non-linear boundaries.
  - B. It regularizes the data to prevent overfitting.
  - C. It is used to calculate probabilities of the predictions.
  - D. It only works for binary classification problems.

Example MCQ

- Which of the following methods is used to evaluate the accuracy of a classification model?
  - A. Mean Squared Error (MSE)
  - B. Cross-entropy Loss
  - C. Confusion Matrix
  - D. Root Mean Squared Error (RMSE)

Example MCQ

- In logistic regression, what does the sigmoid function output represent?
  - A. The actual class label of the data point
  - B. The probability of the data point belonging to a particular class
  - C. The distance of the data point from the decision boundary
  - D. A constant threshold for classification