CS 383: Machine Learning

Prof Adam Poliak

Fall 2024

11/07/2024

Lecture 22

Announcements – Remaining Assignments

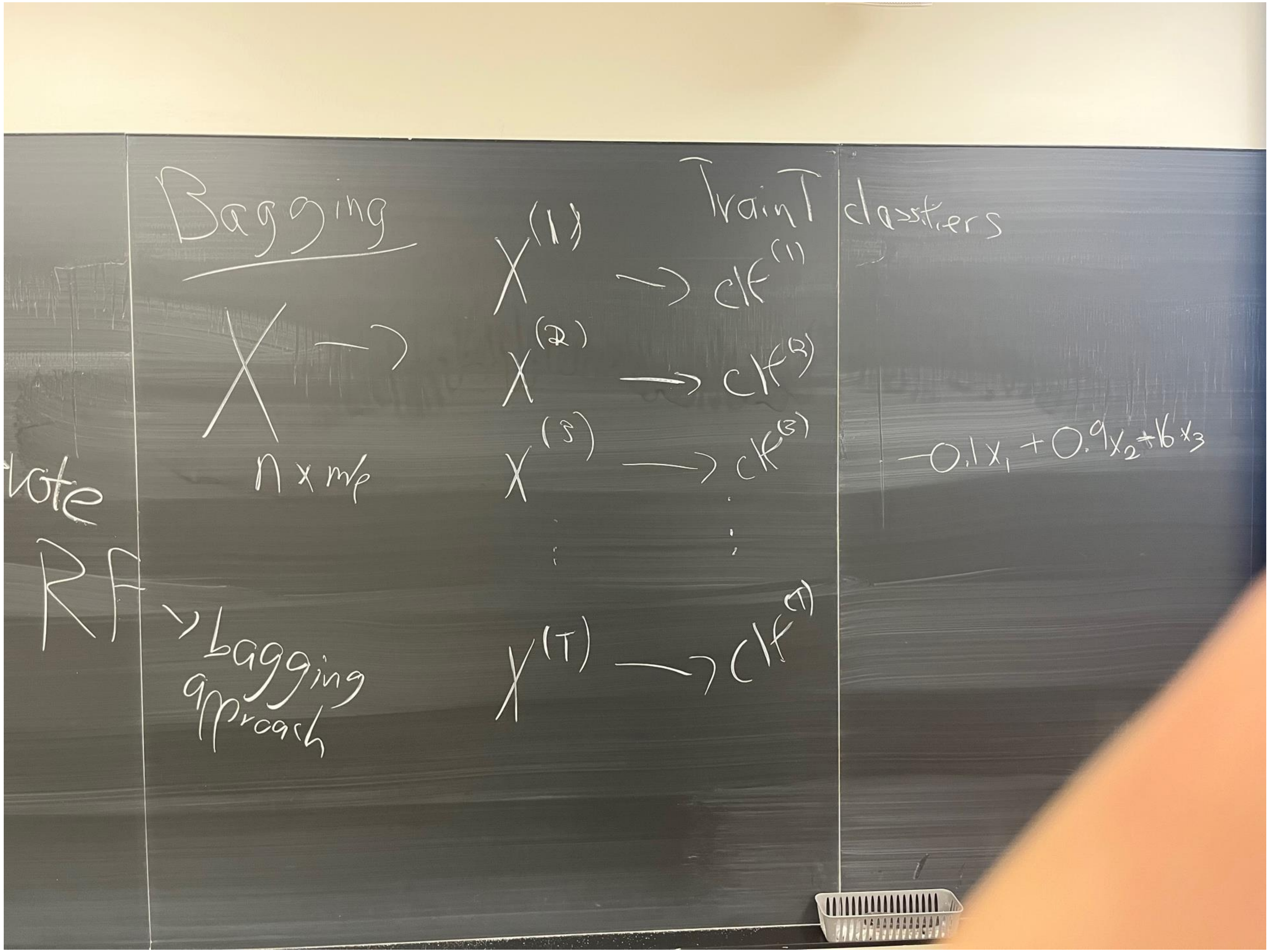
HW06: extending deadline to Friday 11/15

HW07: due Friday 11/22

HW08: due Friday 12/06

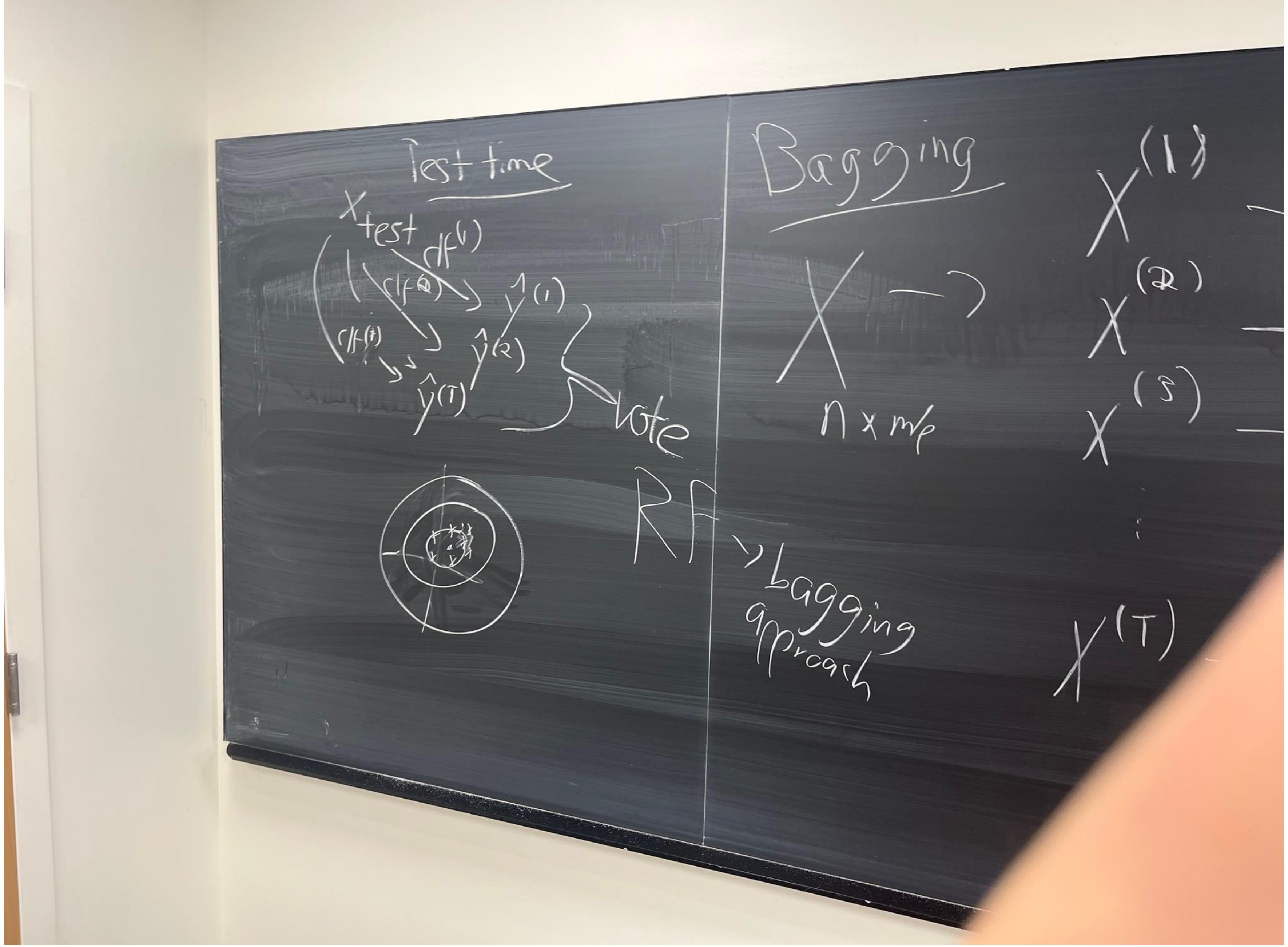
Project Proposal due Thursday 11/14

HW06

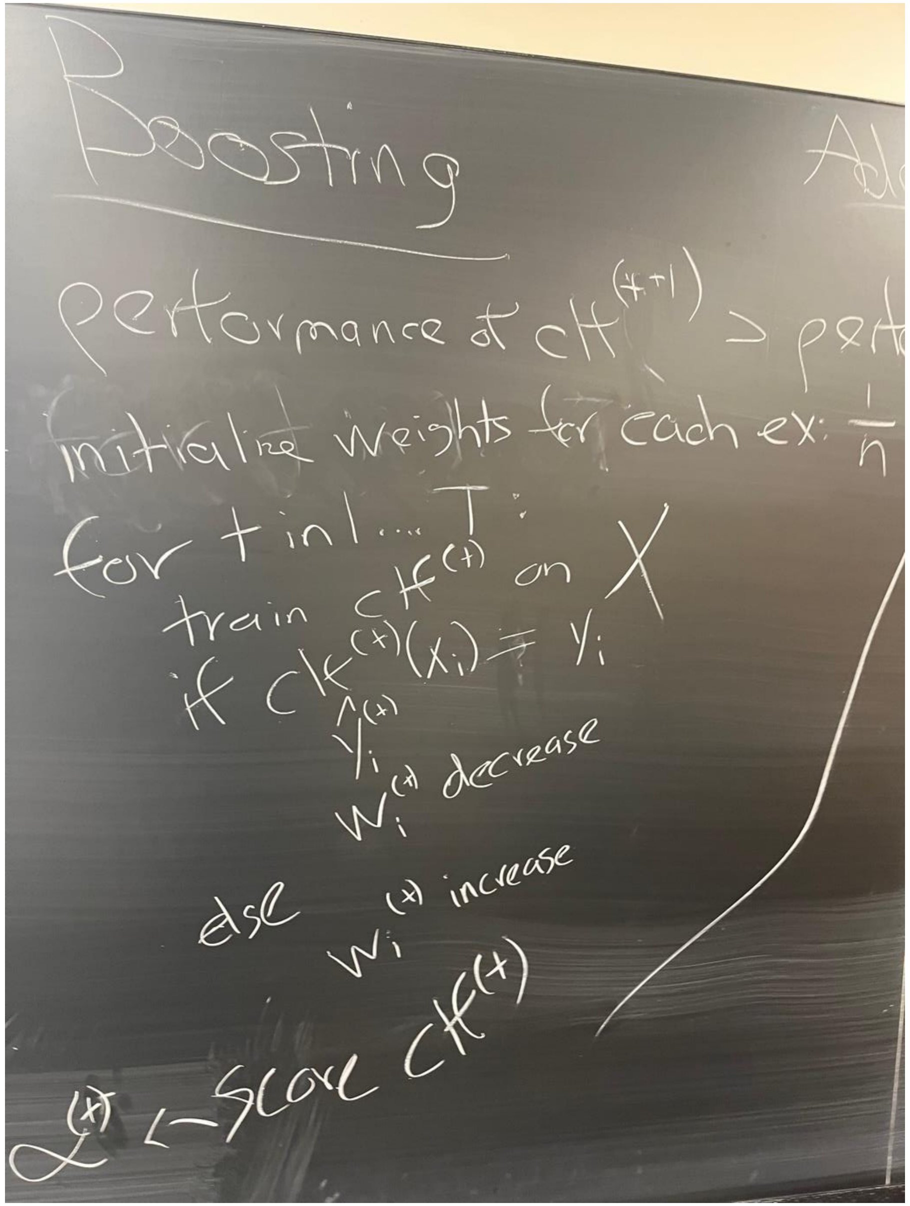


Bagging
n x m/p
Train T classifiers
X^(1) → clf^(1)
X^(2) → clf^(2)
X^(3) → clf^(3)
X^(T) → clf^(T)
RF → bagging approach
-0.1x1 + 0.9x2 + 16x3

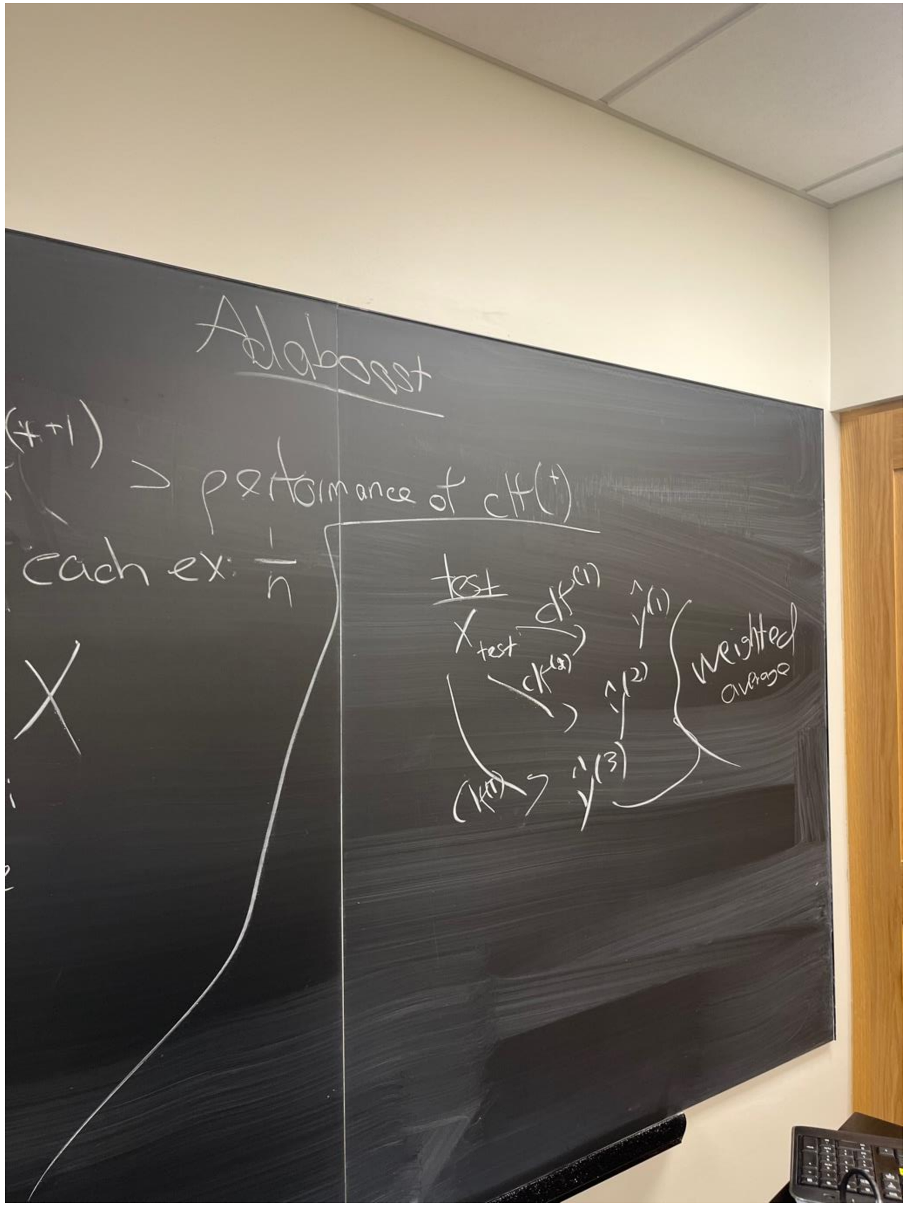
HW06


Test time
X_{test}
$clf^{(1)}$
$clf^{(2)}$
$clf^{(T)}$
$\hat{y}^{(1)}$
$\hat{y}^{(2)}$
$\hat{y}^{(T)}$
vote
RF
Bagging
X
$n \times m/p$
$X^{(1)}$
$X^{(2)}$
$X^{(3)}$
$X^{(T)}$
bagging approach

HW06


Boosting
performance of clf^(t+1) > perf
initialize weights for each ex: 1/n
for t in 1...T:
train clf^(t) on X
if clf^(t)(x_i) = y_i
ŷ_i^(t)
w_i^(t) decrease
else
w_i^(t) increase
α^(t) <- score clf^(t)

HW06


Adaboost
performance of clf(t)
each ex: 1/n
test
weighted
average

Terminology

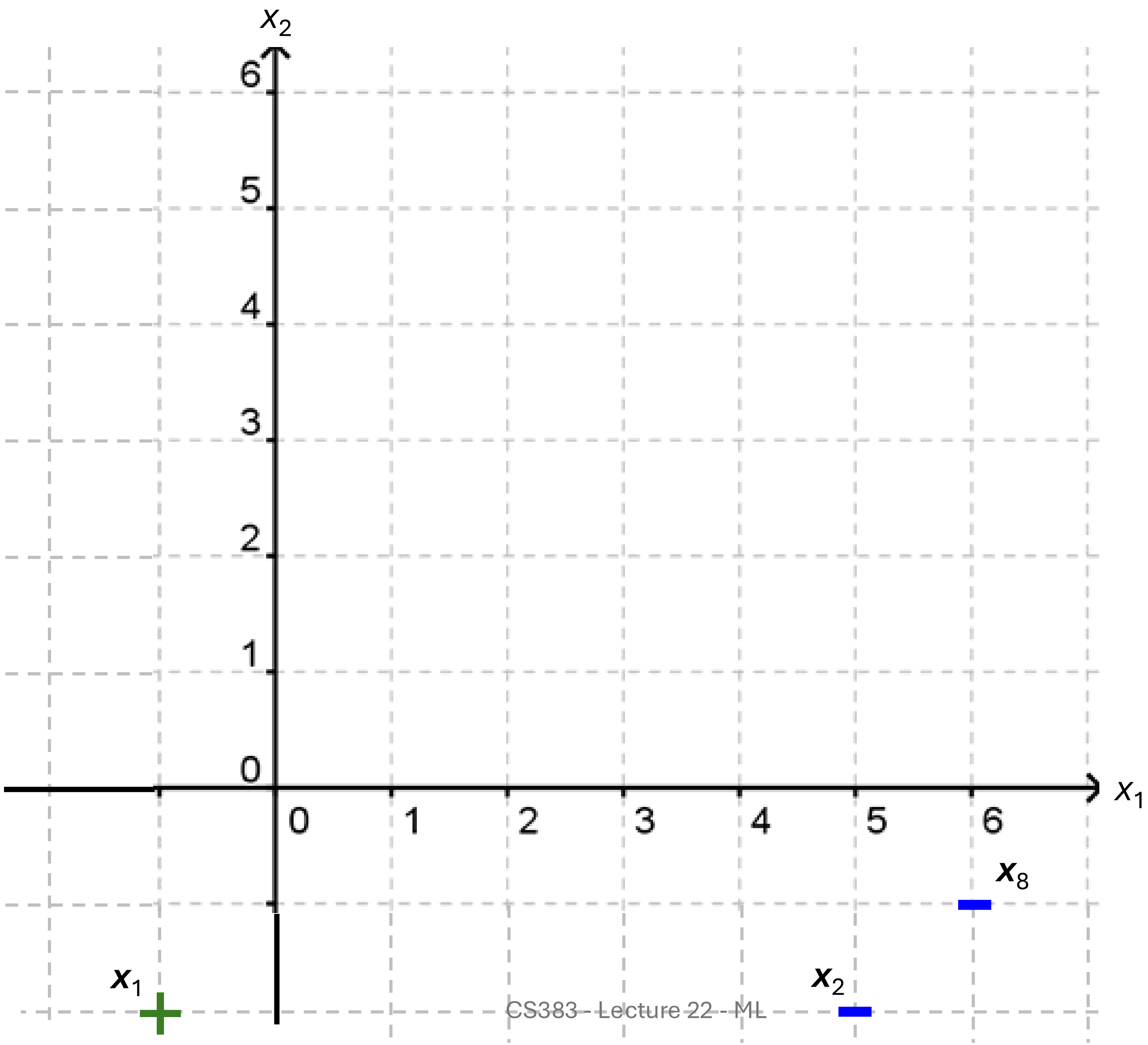
Norm – the length of a vector

L1 (Manhattan distance): $||\vec{v}||_1 = \sum_i |v_i|$

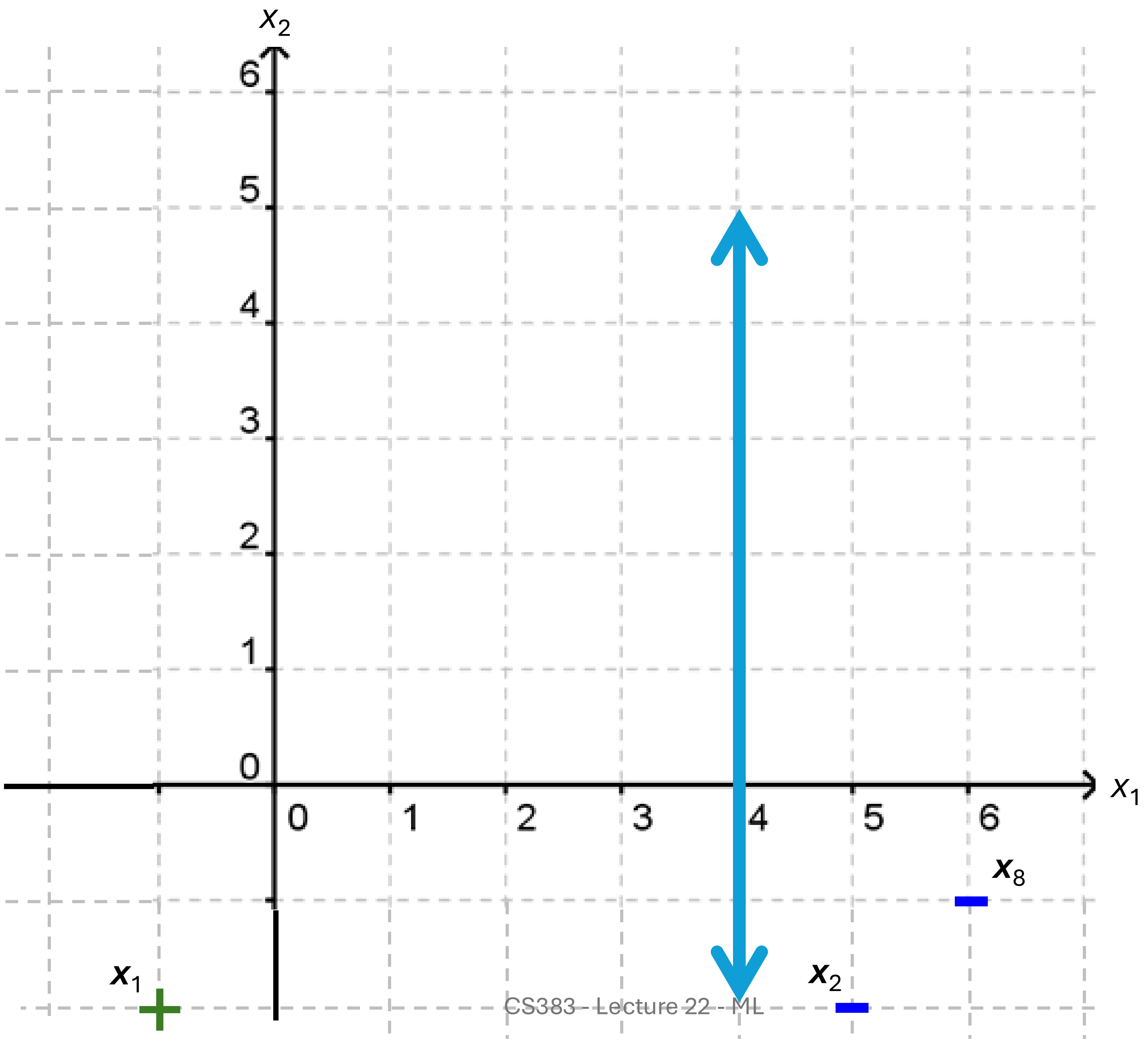
L2 (Euclidian distance): $||\vec{v}||_2 = \sqrt{\sum_i v_i^2}$

Normal vector: vector that is orthogonal/perpendicular to a hyperplane

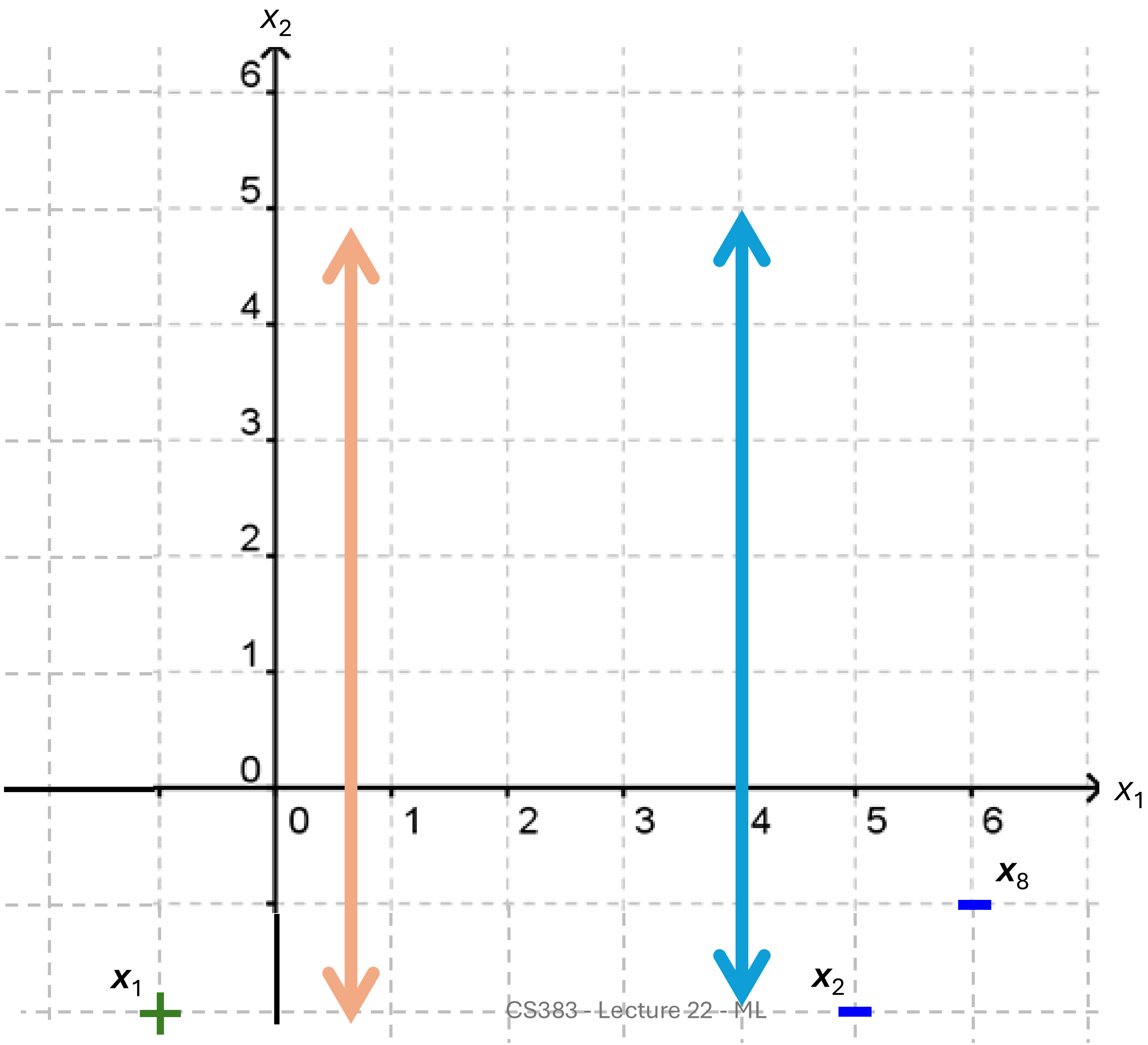
In SVM, $\vec{w} \perp$ hyperplane



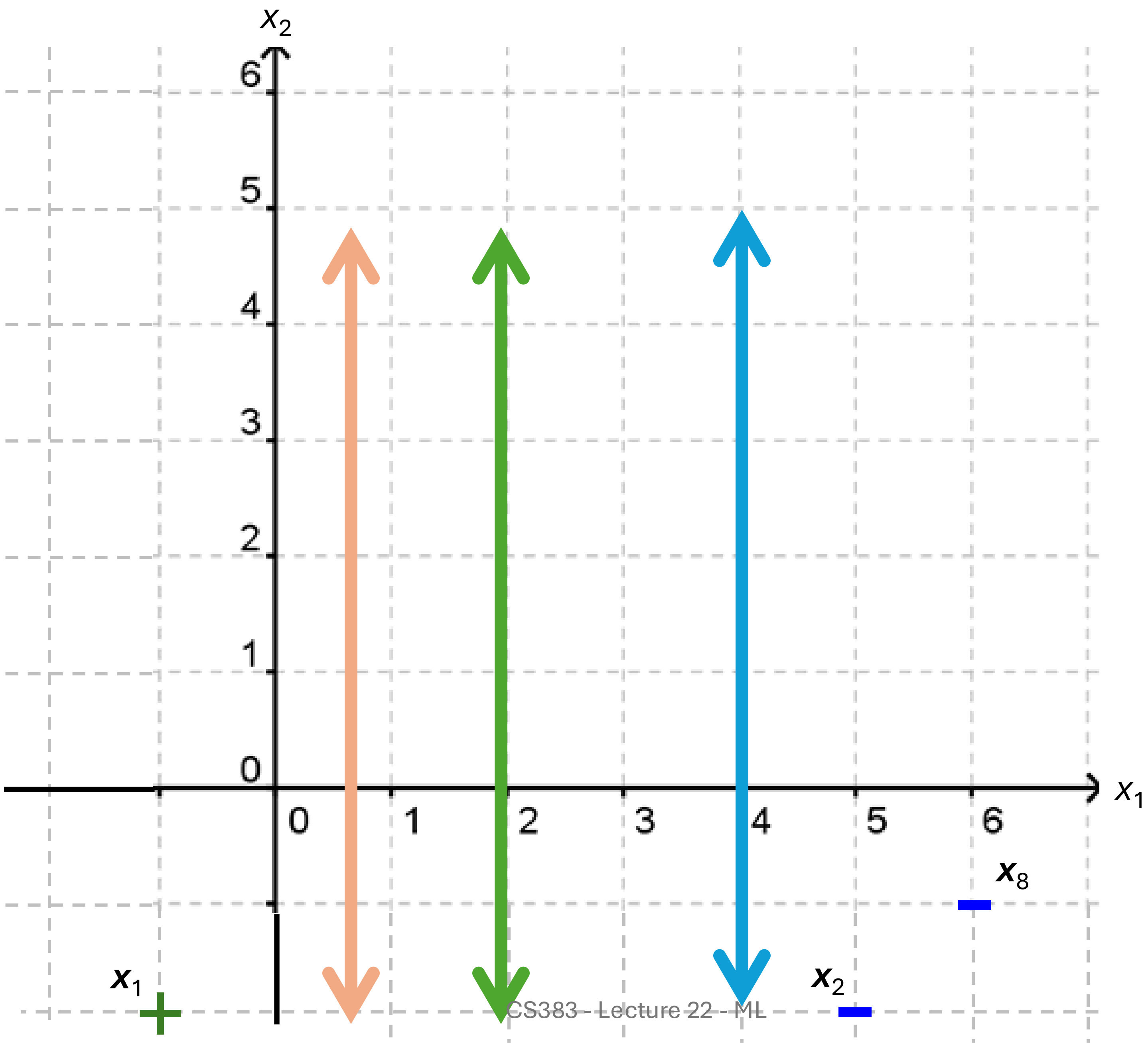
x_2
6
5
4
3
2
1
0
0
1
2
3
4
5
6
x_1
$\boldsymbol{x}_8$
$\boldsymbol{x}_1$
$\boldsymbol{x}_2$



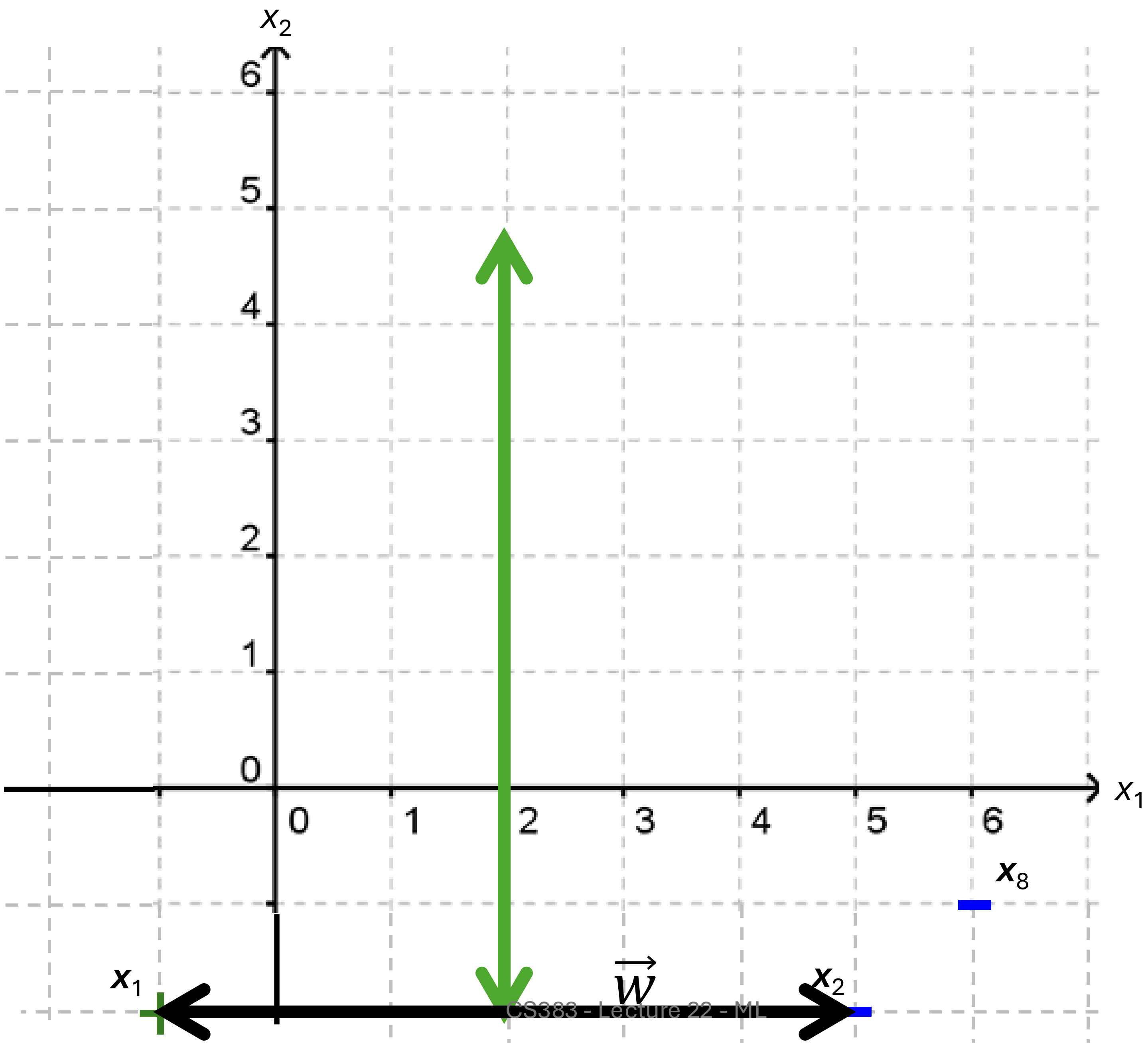
x_2
6
5
4
3
2
1
0
0
1
2
3
4
5
6
x_1
$\boldsymbol{x}_8$
$\boldsymbol{x}_1$
$\boldsymbol{x}_2$

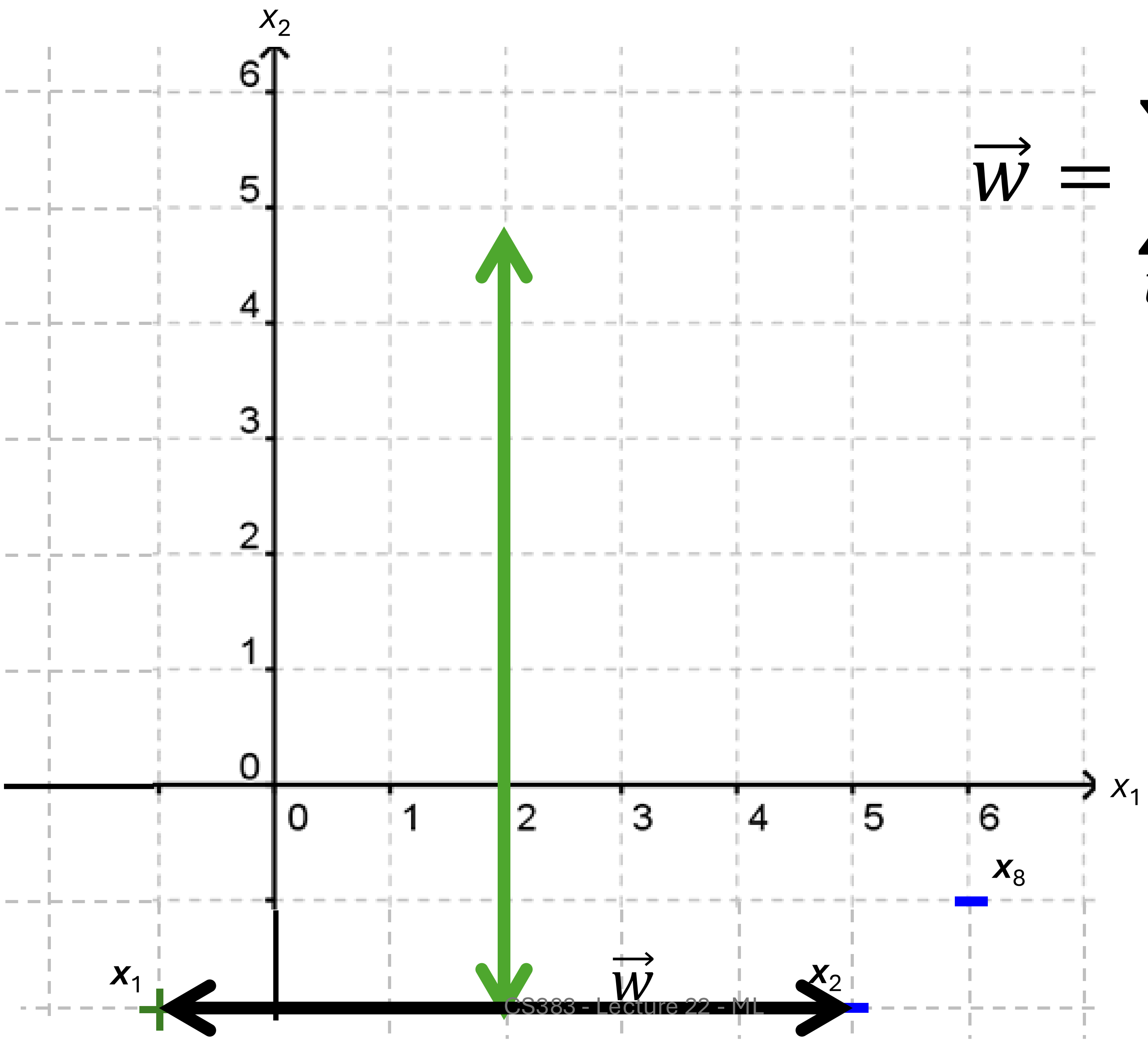


x_2
6
5
4
3
2
1
0
0
1
2
3
4
5
6
x_1
$\boldsymbol{x}_8$
$\boldsymbol{x}_1$
$\boldsymbol{x}_2$

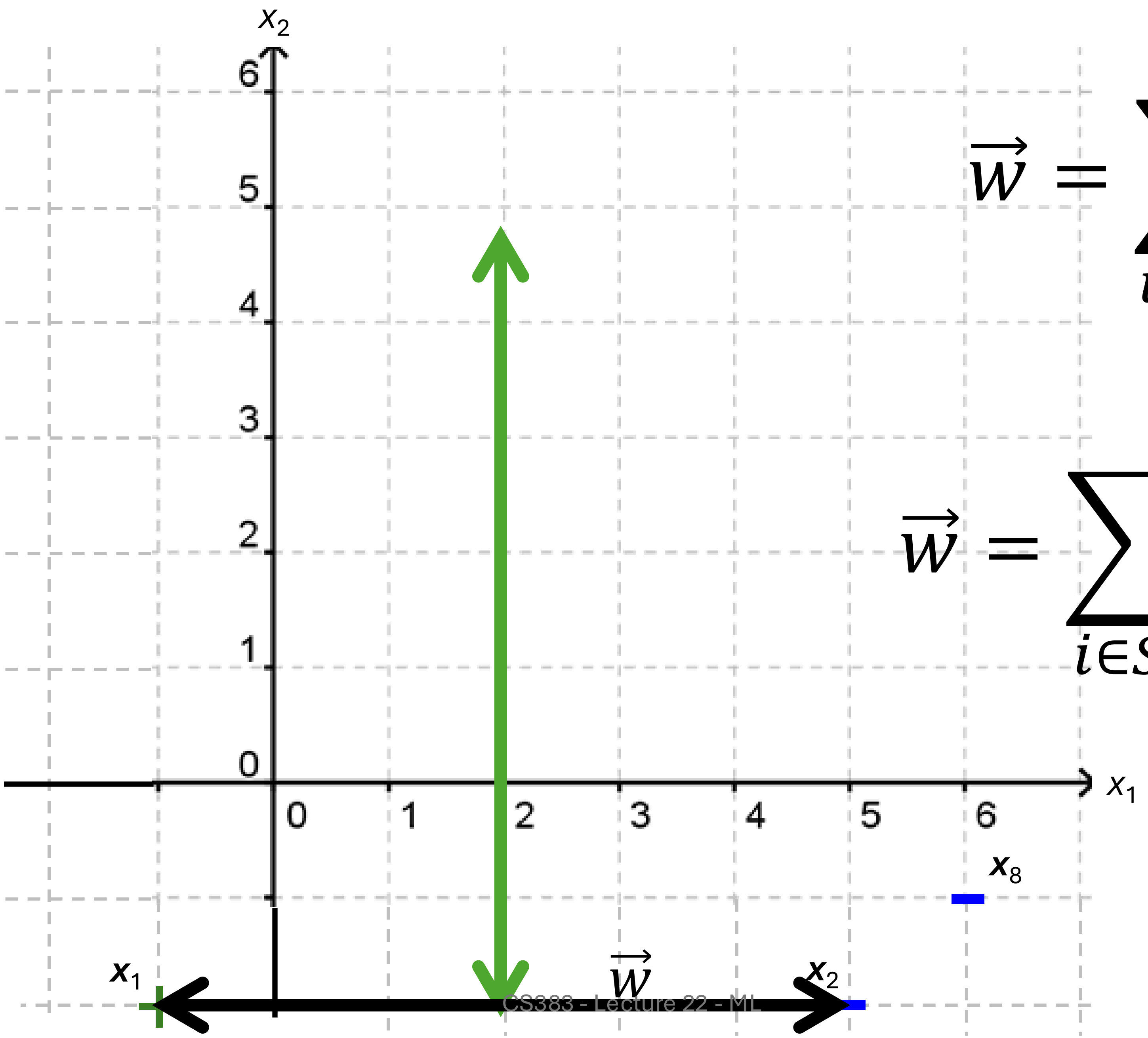


x_2
6
5
4
3
2
1
0
0
1
2
3
4
5
6
x_1
$\boldsymbol{x}_8$
$\boldsymbol{x}_1$
$\boldsymbol{x}_2$


x_2
6
5
4
3
2
1
0
0
1
2
3
4
5
6
x_1
$\boldsymbol{x}_8$
$\boldsymbol{x}_1$
$\vec{w}$
$\boldsymbol{x}_2$



$$\vec{w} = \sum_{i=1}^{n} \alpha_i \, [y_i \, \vec{x_i}]$$



$$\vec{w} = \sum_{i=1}^{n} \alpha_i \left[y_i \, \vec{x_i} \right]$$

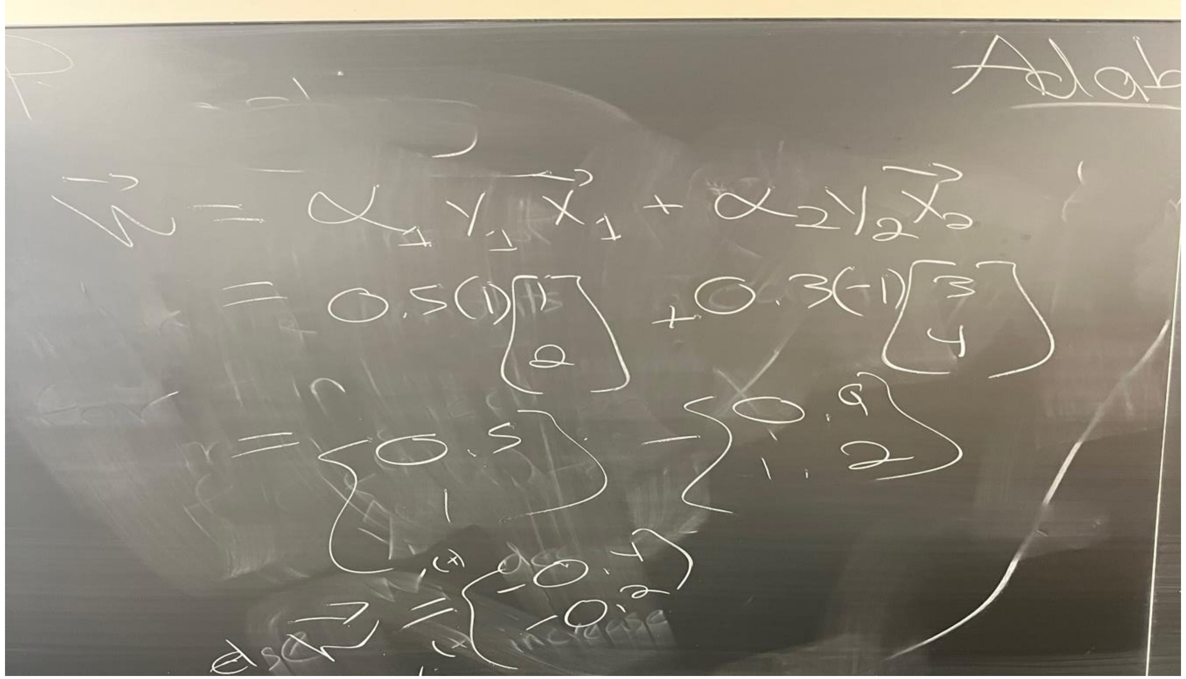
$$\vec{w} = \sum_{i \in S} \alpha_i \left[y_i \, \vec{x_i} \right]$$

Example

Suppose you have two vectors: $x_1 = [1,2], x_2 = [3,4]$ with labels $y_1 = +1, y_2 = -1$, and $\alpha_1 = 0.5, \alpha_2 = 0.3$.

What is $\vec{w}$?

[-0.4, -0.2]



Meta-optimization process

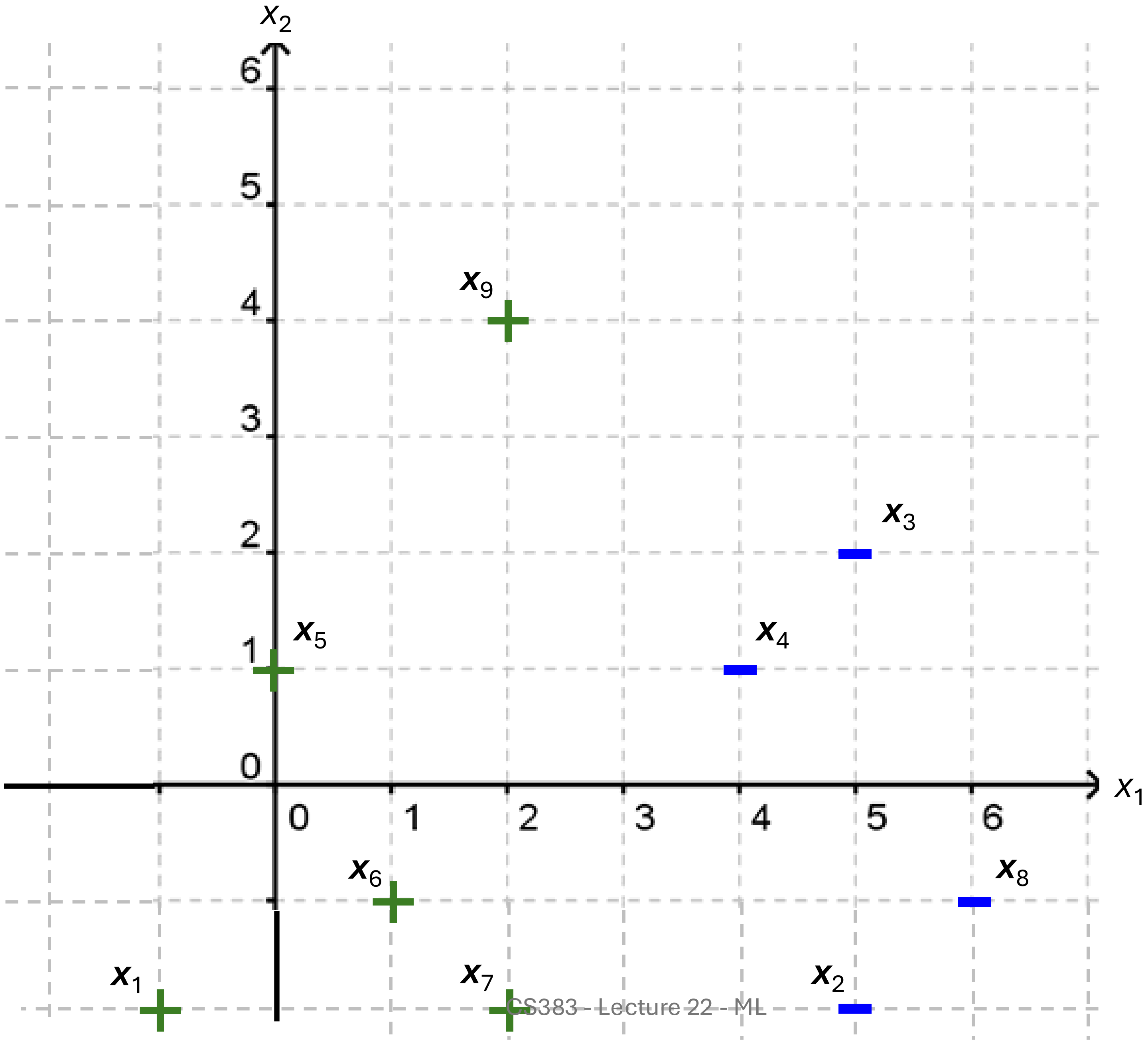
- Incremental SVM optimization algorithm

- Choose a subset S of examples and run optimization to get alpha values

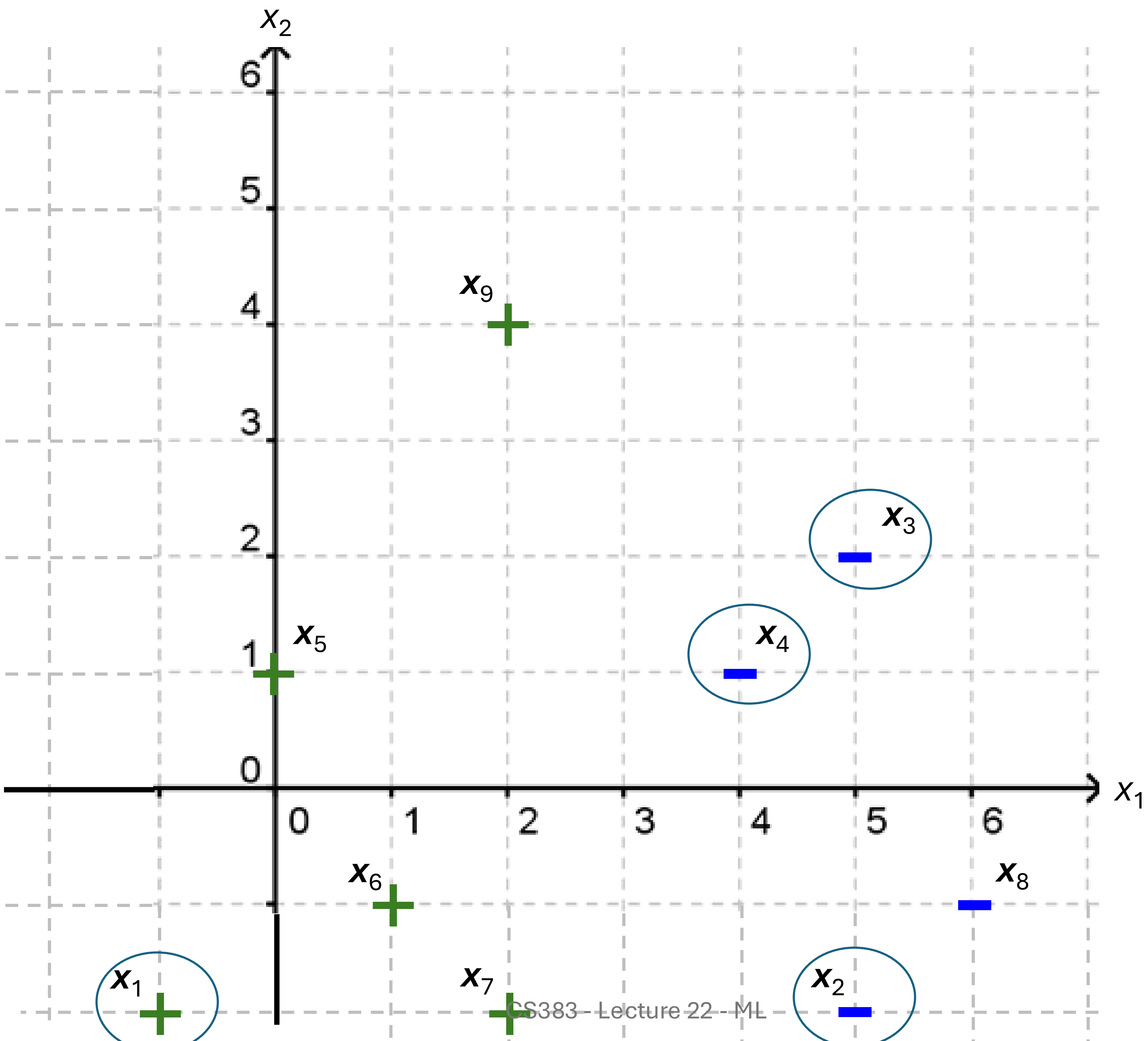
- Identify which alpha values are 0 => these cannot be support vectors in final solution!

- Discard these points and add new ones; repeat

Meta-optimization: example



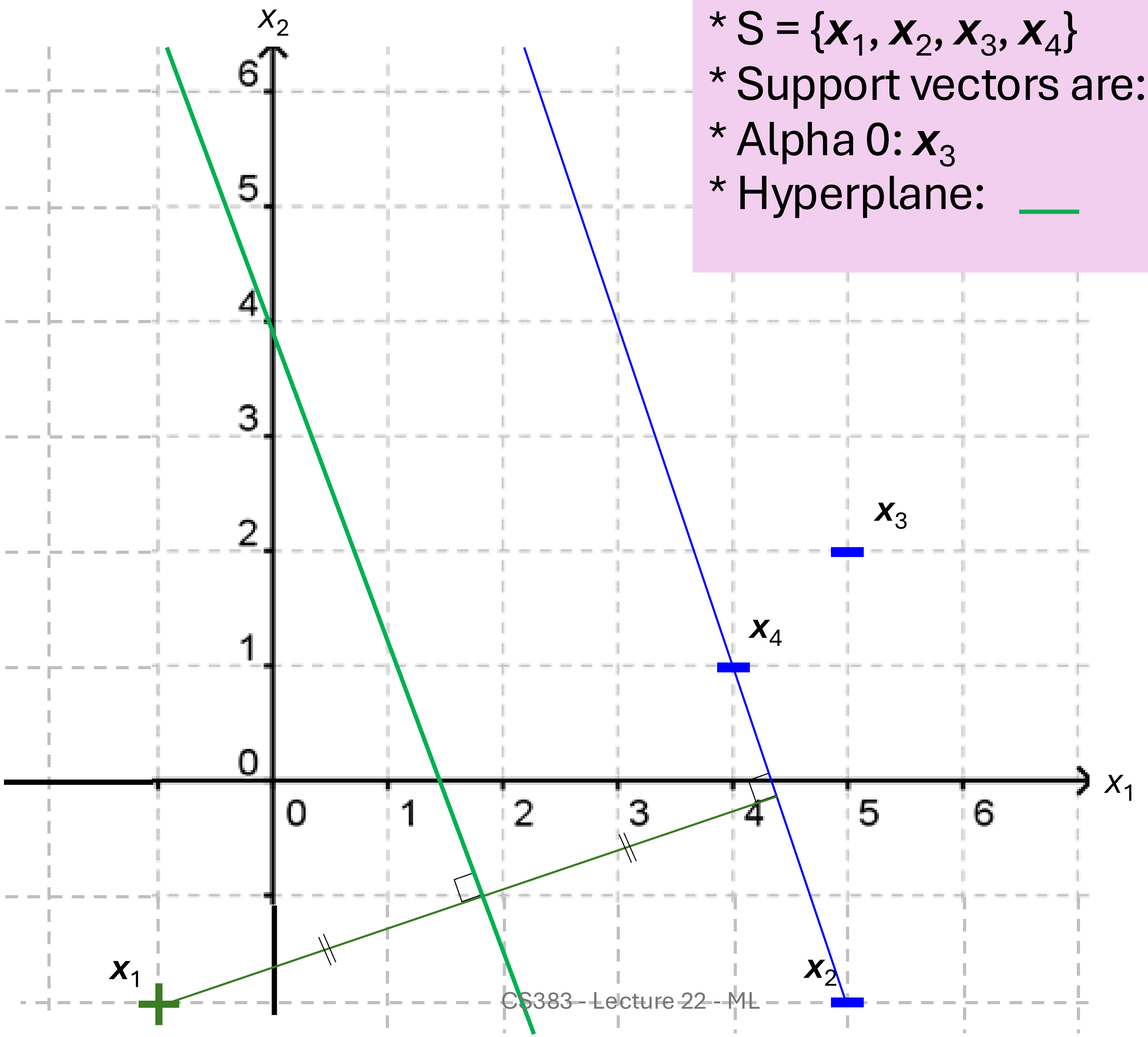
Meta-optimization: example



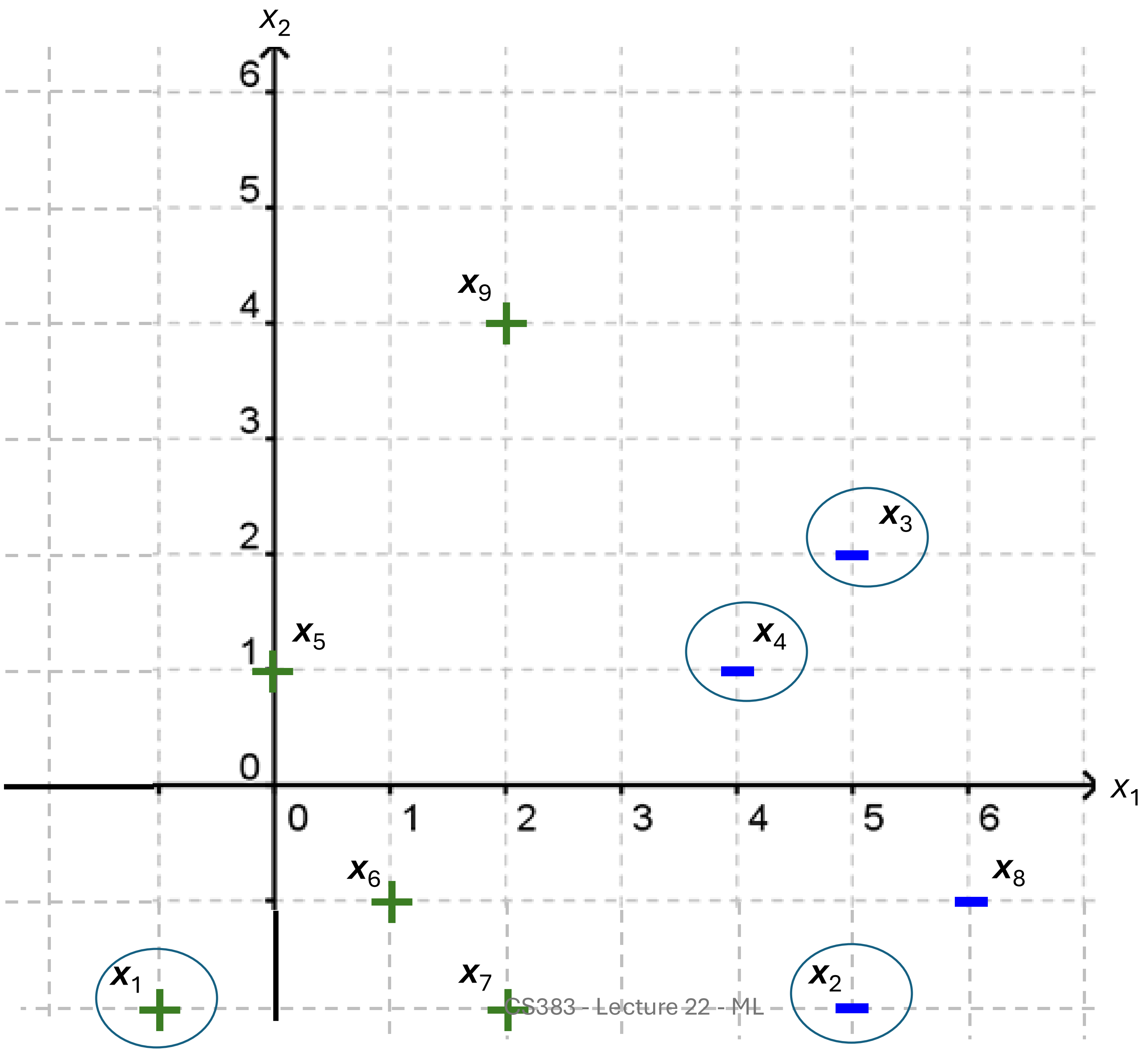
$K = 4$

Round 1:

* S = {$\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4$}
* Support vectors are: $\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4$
* Alpha 0: $\boldsymbol{x}_3$
* Hyperplane: ___



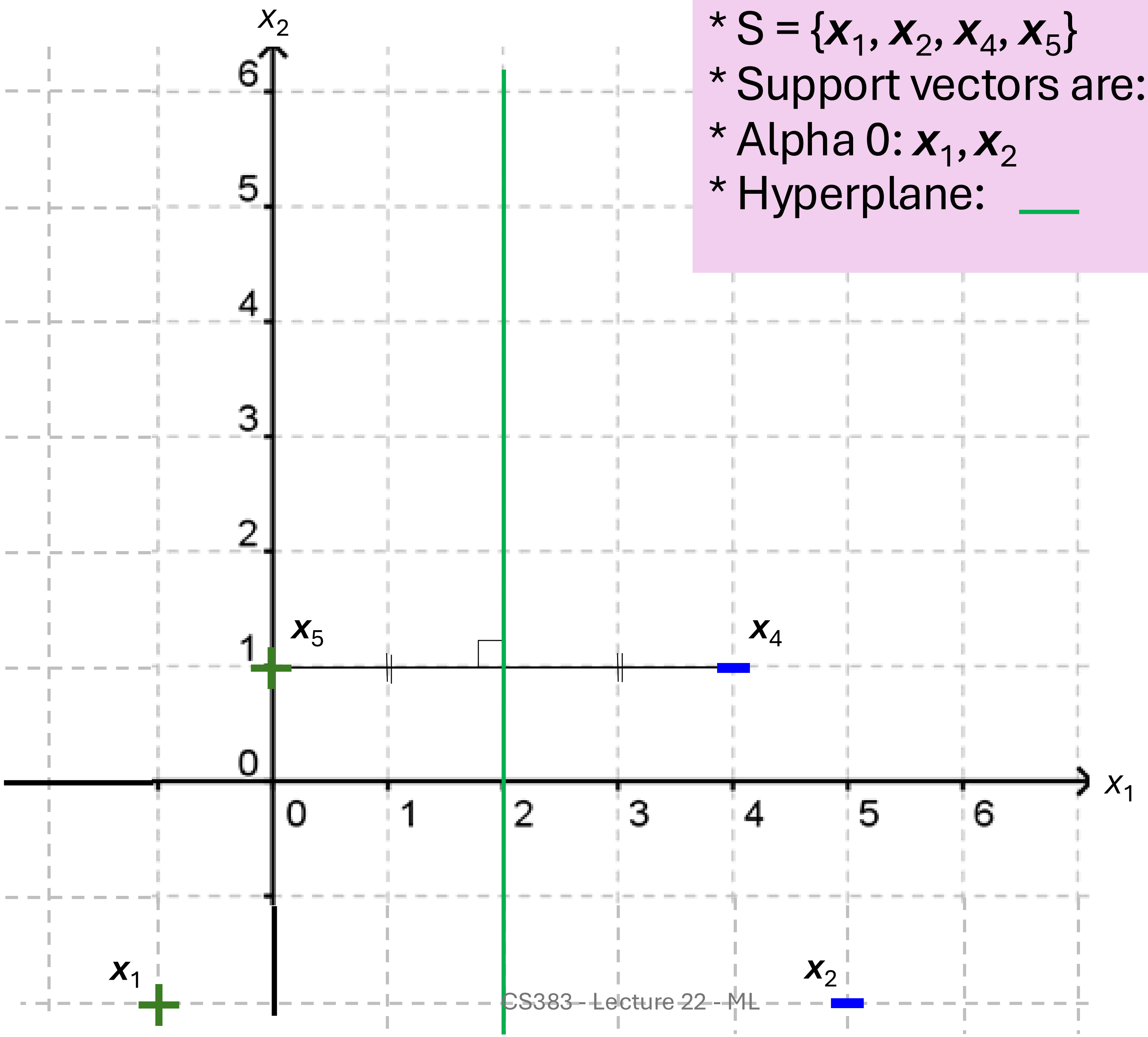
Meta-optimization: example



$K = 4$

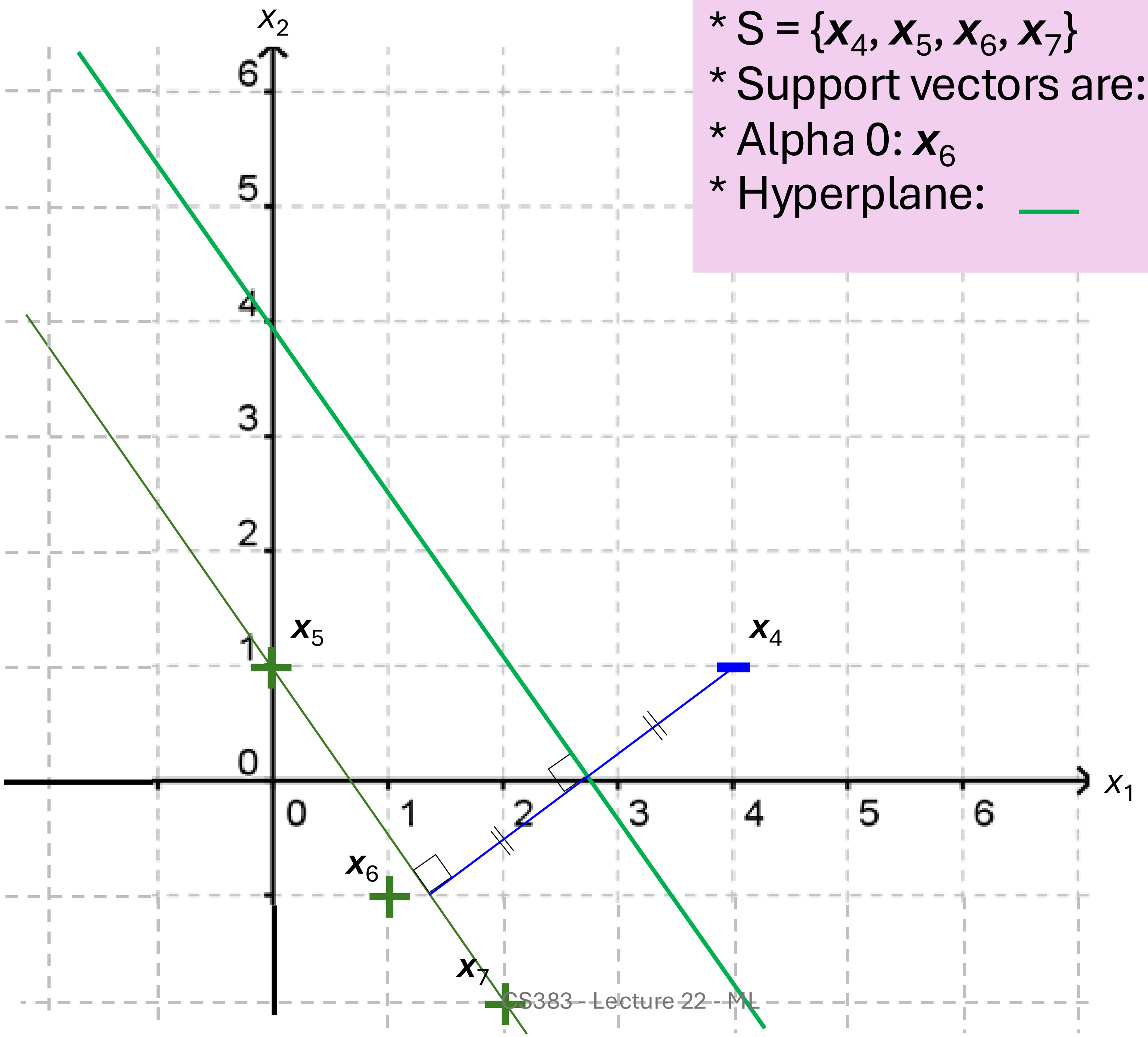
Round 1:

* S = {$\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_4, \boldsymbol{x}_5$}
* Support vectors are: $\boldsymbol{x}_4, \boldsymbol{x}_5$
* Alpha 0: $\boldsymbol{x}_1, \boldsymbol{x}_2$
* Hyperplane: ___



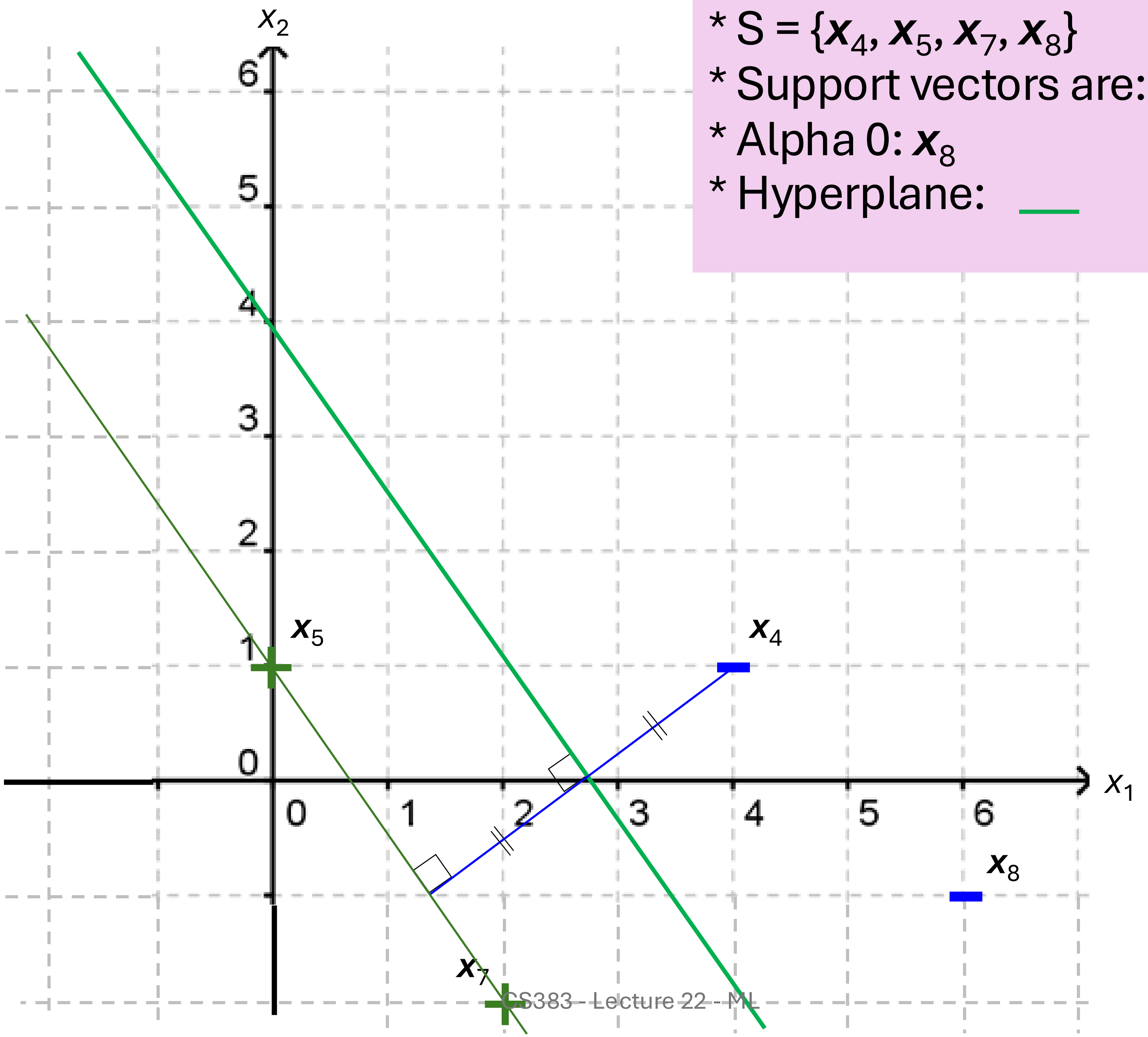
Round 3:

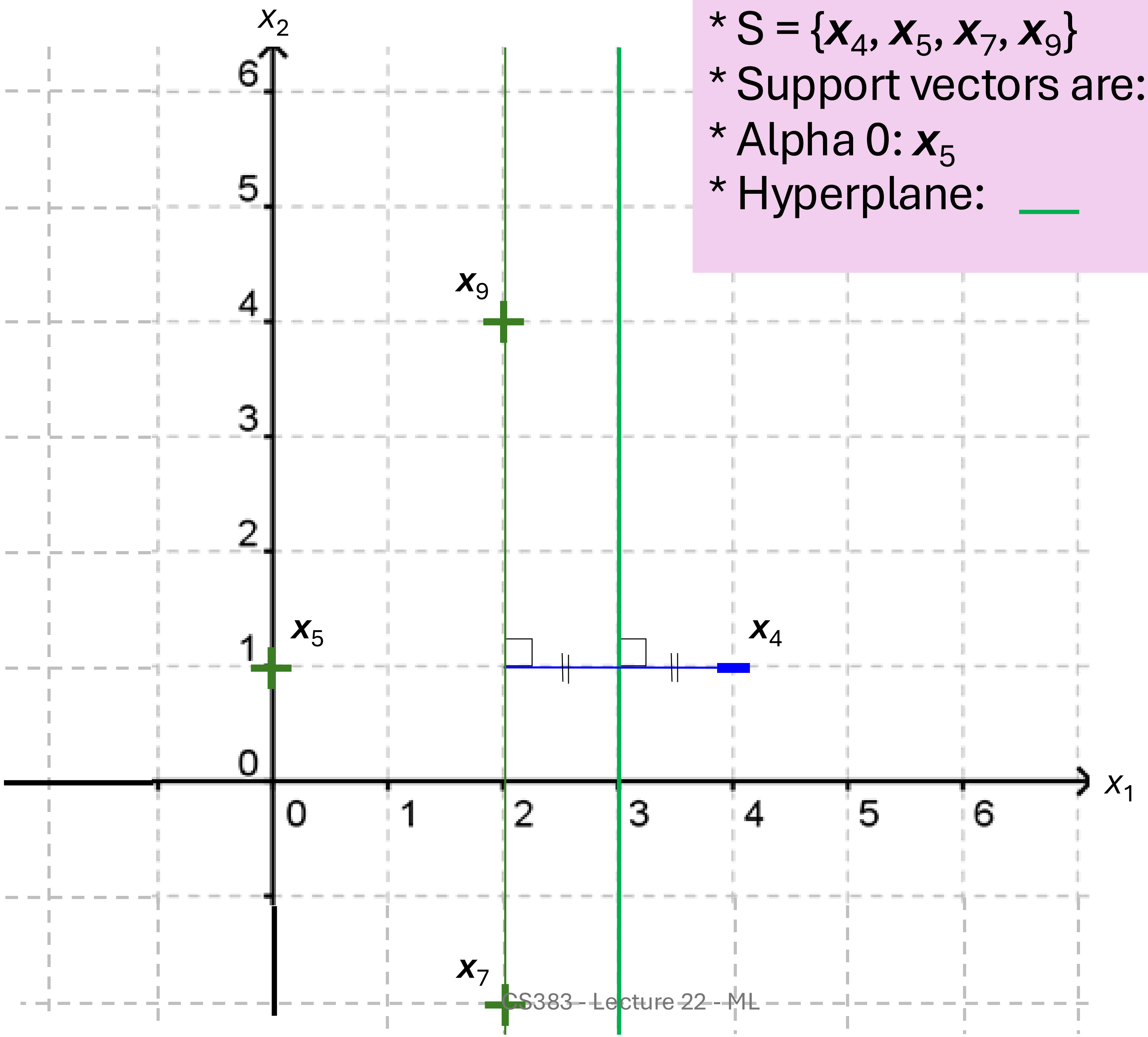
* S = {$\boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_6, \boldsymbol{x}_7$}
* Support vectors are: $\boldsymbol{x}_4, \boldsymbol{x}_5, \boldsymbol{x}_7$
* Alpha 0: $\boldsymbol{x}_6$
* Hyperplane:



Round 4:

* S = {$\boldsymbol{x}_4$, $\boldsymbol{x}_5$, $\boldsymbol{x}_7$, $\boldsymbol{x}_8$}
* Support vectors are: $\boldsymbol{x}_4$, $\boldsymbol{x}_5$, $\boldsymbol{x}_7$
* Alpha 0: $\boldsymbol{x}_8$
* Hyperplane:

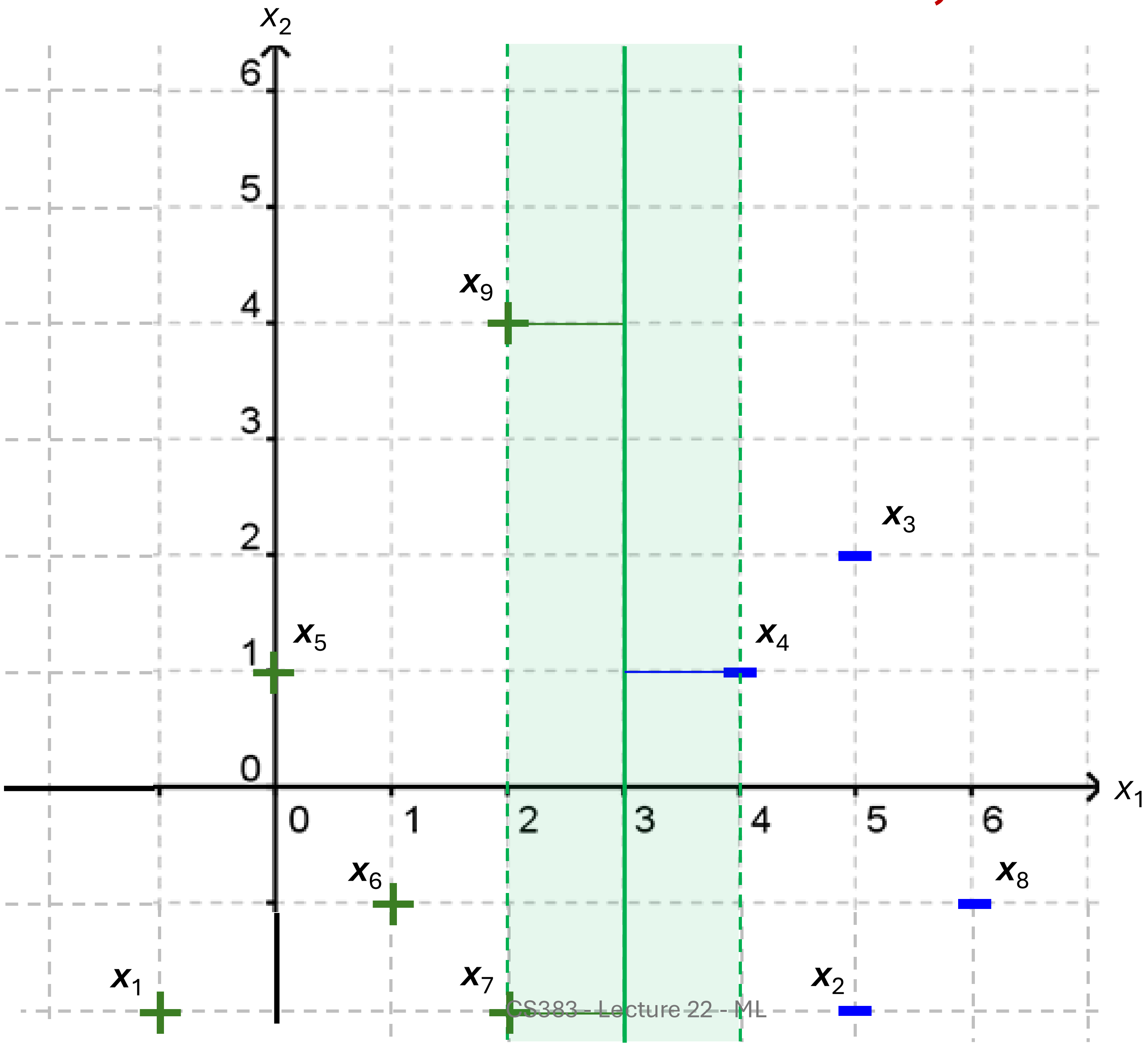




Round 5:

* S = {$\boldsymbol{x}_4$, $\boldsymbol{x}_5$, $\boldsymbol{x}_7$, $\boldsymbol{x}_9$}
* Support vectors are: $\boldsymbol{x}_4$, $\boldsymbol{x}_7$, $\boldsymbol{x}_9$
* Alpha 0: $\boldsymbol{x}_5$
* Hyperplane:

Handout 18, Final Solution



Lagrange Multiplier for SVM

$$\min_{\vec{w},b} \quad \frac{1}{2}\|\vec{w}\|^2$$

$$\text{s.t.} \quad y_i(\vec{w} \cdot \vec{x}_i + b) \geq 1, \quad i = 1, \cdots, n$$

$$h(\overrightarrow{w}, b, \vec{\alpha}) = \frac{1}{2}\left||\overrightarrow{w}|\right|^2 - \sum_{i=1}^{n} \alpha_i \left[y_i \overrightarrow{(w}\ \overrightarrow{x_i})\right]$$

Dual form

$$\max W(\vec{\alpha}) = \sum_{i}^{n} \alpha_i - \frac{1}{2} \sum_{i}^{n} \sum_{j}^{n} y_i y_j \alpha_i \alpha_j \boxed{\vec{x_i}\,\vec{x_j}}$$

$$s.t.\, \alpha_i > 0\ \forall i\ \&\ \sum_{i}^{n} \alpha_i y_i = 0$$

Kernel Idea

- By solving the dual form of the problem, we have seen how all computations can be done in terms of inner products between examples

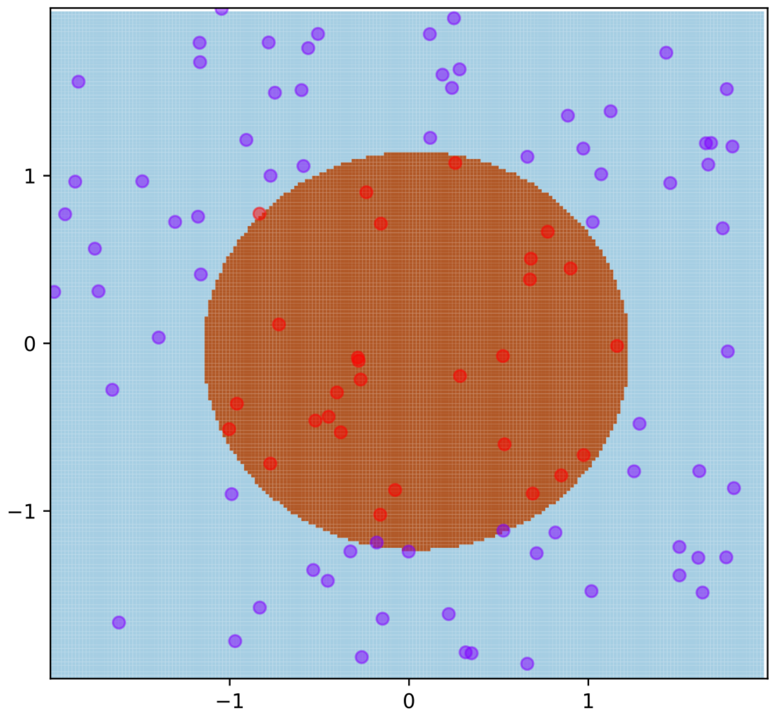
- One example of an inner product is the dot product, which is the linear version of SVMs

- But there are many others!

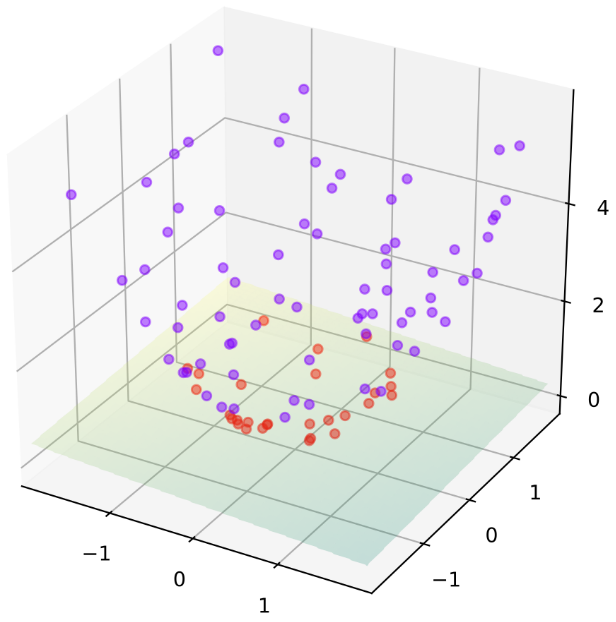
- Intuition: if points are close together, their kernel function will have a large value (measure of similarity)

Kernel Trick example

Feature mapping: $\varphi(\boldsymbol{x}) = (x_1,\ x_2,\ x_1^2 + x_2^2)$



Original feature space



Mapping after applying kernel (can now find a hyperplane)

Kernel function: $K(\boldsymbol{x},\ \boldsymbol{z}) = \boldsymbol{x} \bullet \boldsymbol{z} + ||\boldsymbol{x}||^2\ ||\boldsymbol{z}||^2$

Image: Shiyu Ji (wikipedia)

Gaussian Kernel

- Gaussian kernel is near 0 when points are far apart and near 1 when they are similar
- Also called Radial Basis Function (RBF) kernel

$$K(\vec{x}, \vec{z}) = \exp\left(-\frac{\|\vec{x} - \vec{z}\|^2}{2\sigma^2}\right)$$

Gaussian Kernel

- Gaussian kernel is near 0 when points are far apart and near 1 when they are similar
- Also called Radial Basis Function (RBF) kernel

$$K(\vec{x}, \vec{z}) = \exp\left(-\frac{\|\vec{x} - \vec{z}\|^2}{2\sigma^2}\right)$$

Often re-parametrized by gamma (different gamma!)

$$\boxed{\gamma = \frac{1}{2\sigma^2}}$$

$$K(\vec{x}, \vec{z}) = \exp\left(-\gamma\|\vec{x} - \vec{z}\|^2\right)$$

We will tune gamma as part of Homework 7

Soft-margin SVMs (non-separable case)

- Idea: we will use regularization to add a cost for each point being incorrectly classified by the hyperplane

- Hopefully many costs will be 0, but we can accommodate a few outliers

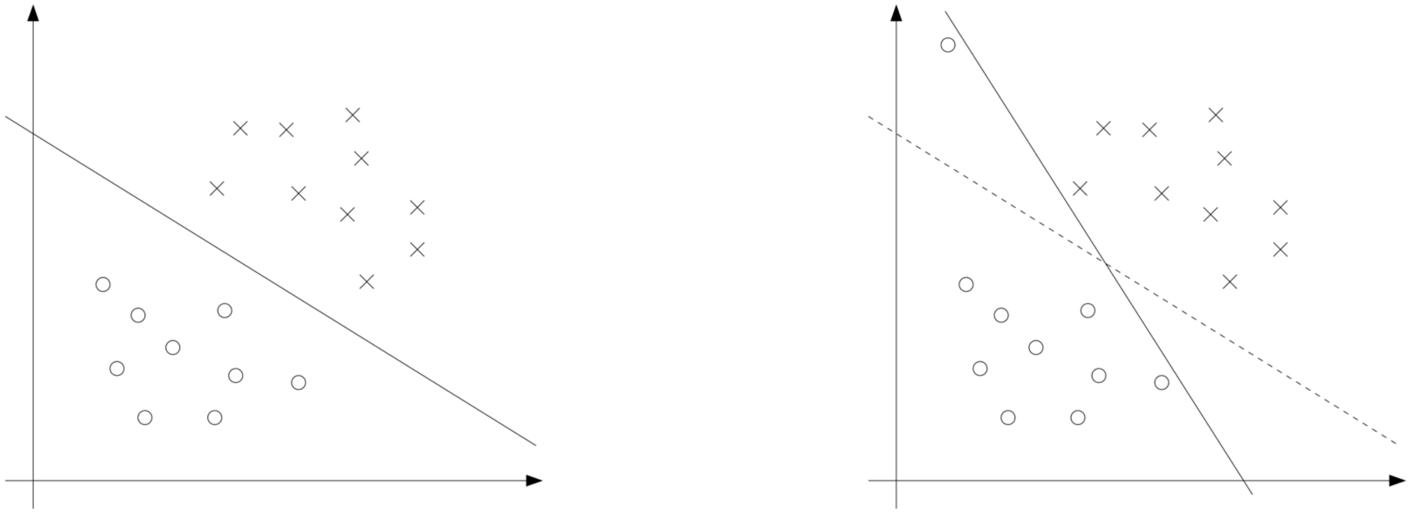


Figure: Andrew Ng

Soft-margin SVMs (non-separable case)

- New optimization problem with regularization
- We will tune the *C* parameter as part of Homework 7

$$\min_{\xi,\vec{w},b} \quad \frac{1}{2}\|\vec{w}\|^2 + C\sum_{i=1}^{n}\xi_i$$

$$\text{s.t.} \quad y_i(\vec{w}\cdot\vec{x}_i + b) \geq 1 - \xi_i, \quad i = 1,\cdots,n$$

$$\text{and} \quad \xi_i \geq 0, \quad i = 1,\cdots,n$$

"flexible margin"