Introduction to Machine Learning

[Fall 2022]

Support Vector Machines (Part 2)

September 27, 2022

Lerrel Pinto

Topics for today

- Whiteboarding SVMs
- Soft margin SVM

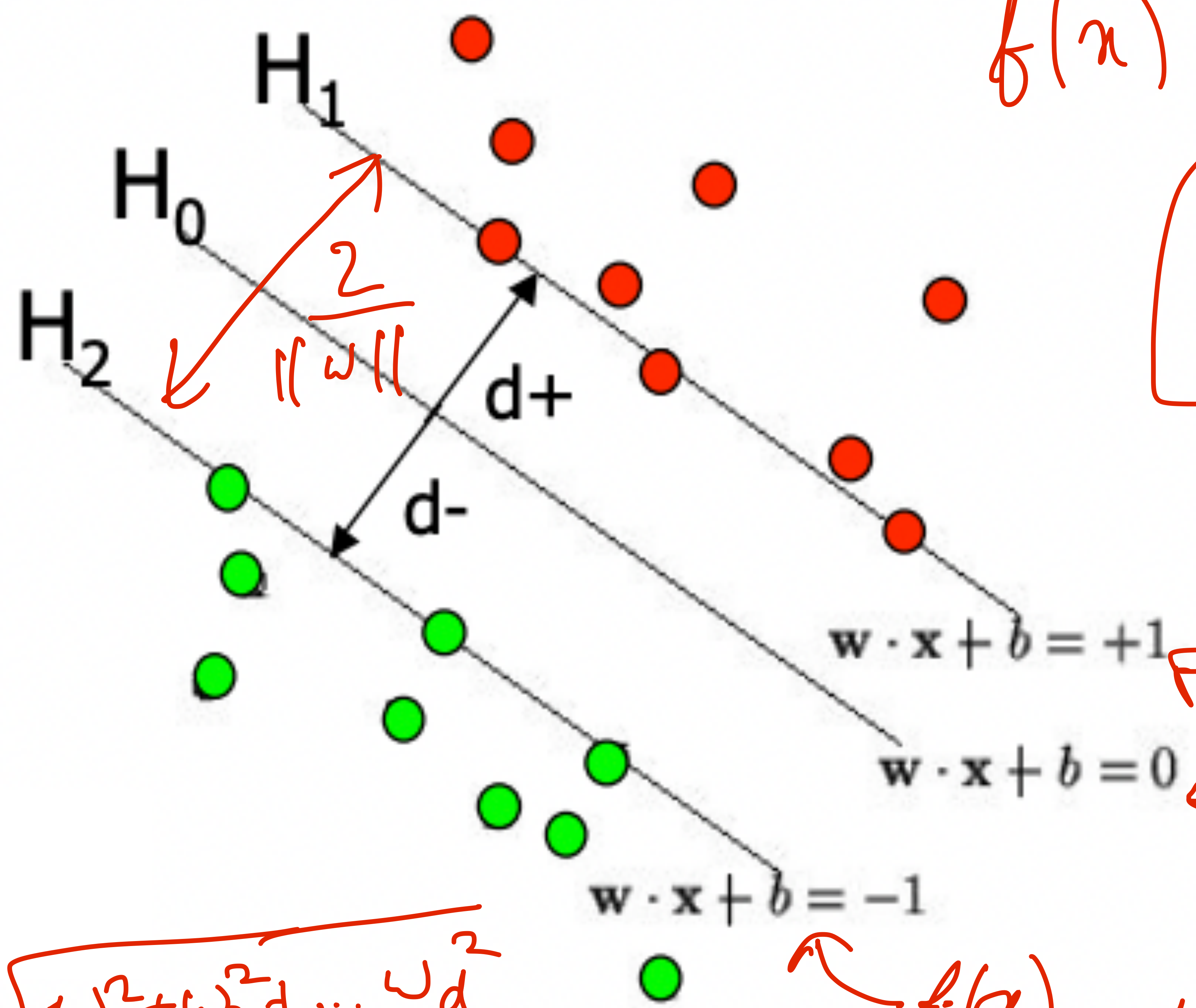
Recap: SVMs

$f(x) := w^T x + b$

$f(x) \geq 1 \rightarrow$ class #1

$f(x) \leq -1 \rightarrow$ class #2

$\boxed{f(x) = 0}$



$\frac{2}{\|w\|}$

$f(x) = 1$

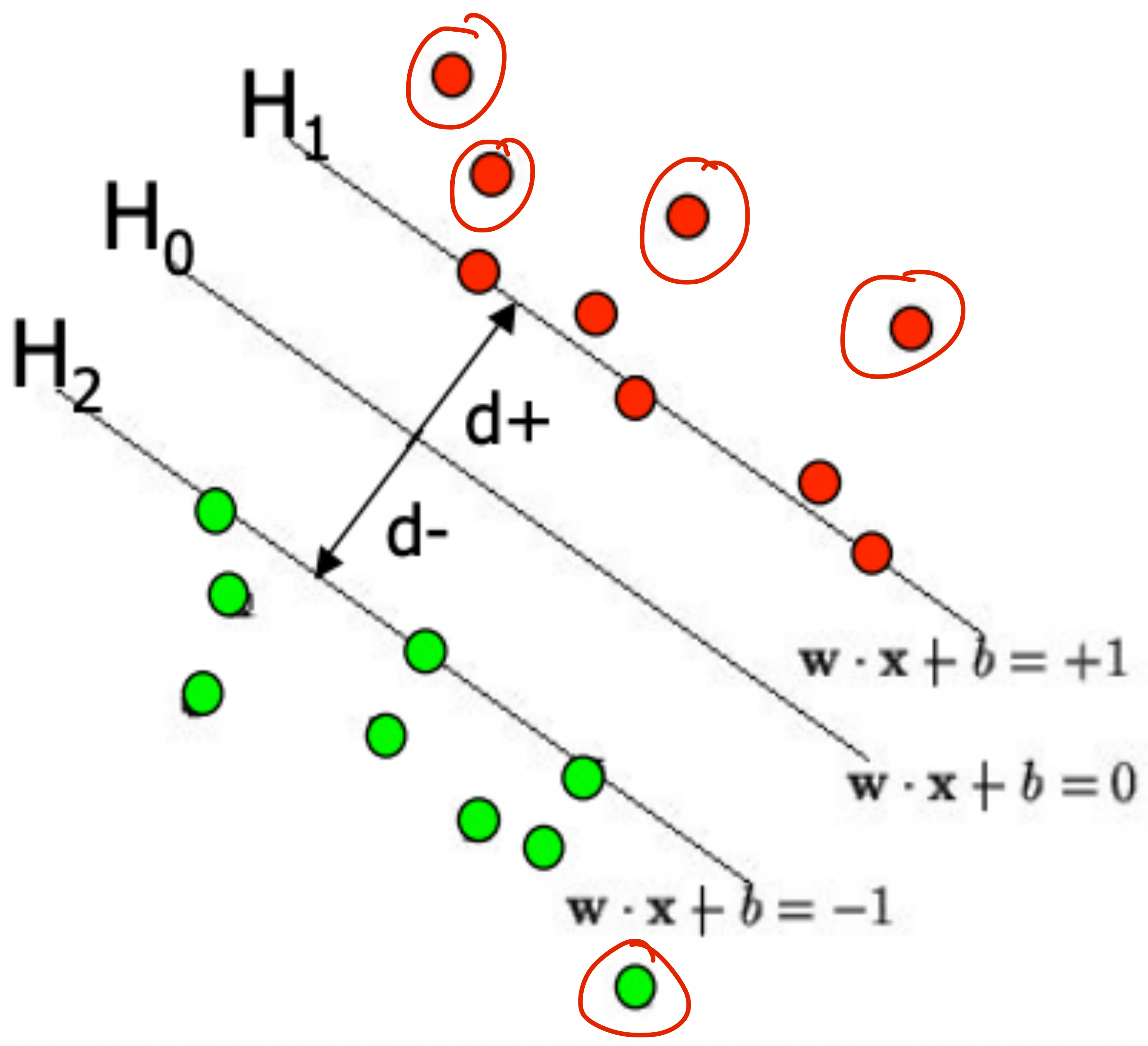
$f(x) = -1$

$\sqrt{w_1^2 + w_2^2 + \dots w_d^2}$

$2\|w\|$ $\quad$ $\|w\|^2$

Credits: R. Berwick (https://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf)

Recap: SVMs



- Goal: Maximize margin / Minimize $||w||^2$
  - Also need to satisfy $y^i f(x^i) \geq 1$ for all datapoints (x^i, y^i).

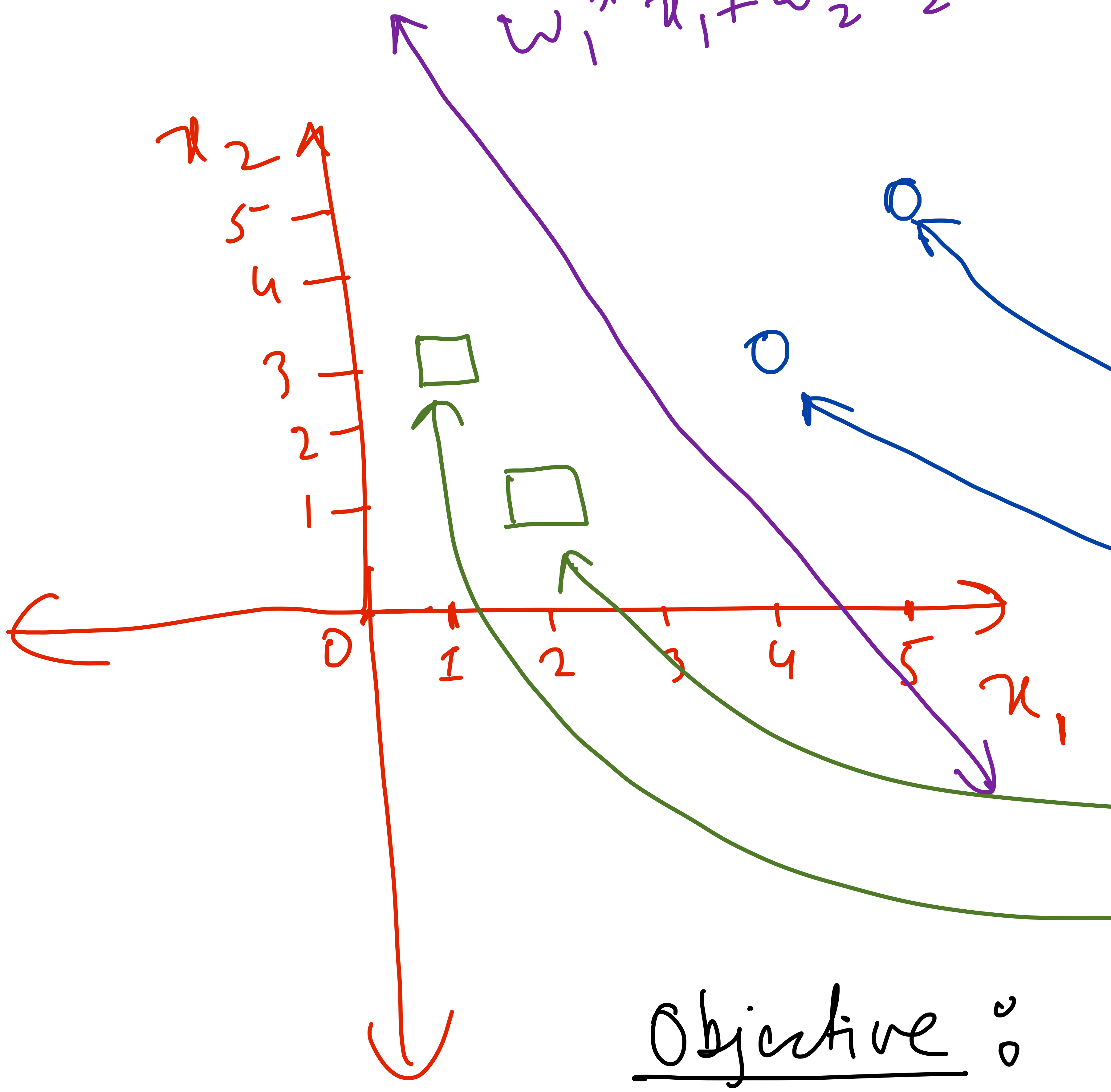
$$\boxed{\min_w ||w||^2 \text{ subject to } y^i(w^T x^i + b) \geq 1}$$

- Can be solved as a quadratic optimization problem with linear constraints.

Credits: R. Berwick (https://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf)

Example 1

$w_1^* x_1 + w_2^* x_2 + b^* = 0$



Data

$y^i = +1$ ○ $\left(\begin{bmatrix} 4 \\ 3 \end{bmatrix}, \begin{bmatrix} 5 \\ 5 \end{bmatrix} \right)$

$y^i = -1$ □ $\left(\begin{bmatrix} 2 \\ 1 \end{bmatrix}, \begin{bmatrix} 1 \\ 3 \end{bmatrix} \right)$

Objective: Optimize (w_1, w_2, b)

$y^i (w_1 x_1^i + w_2 x_2^i + b) \geq 1$

minimize $w_1^2 + w_2^2$

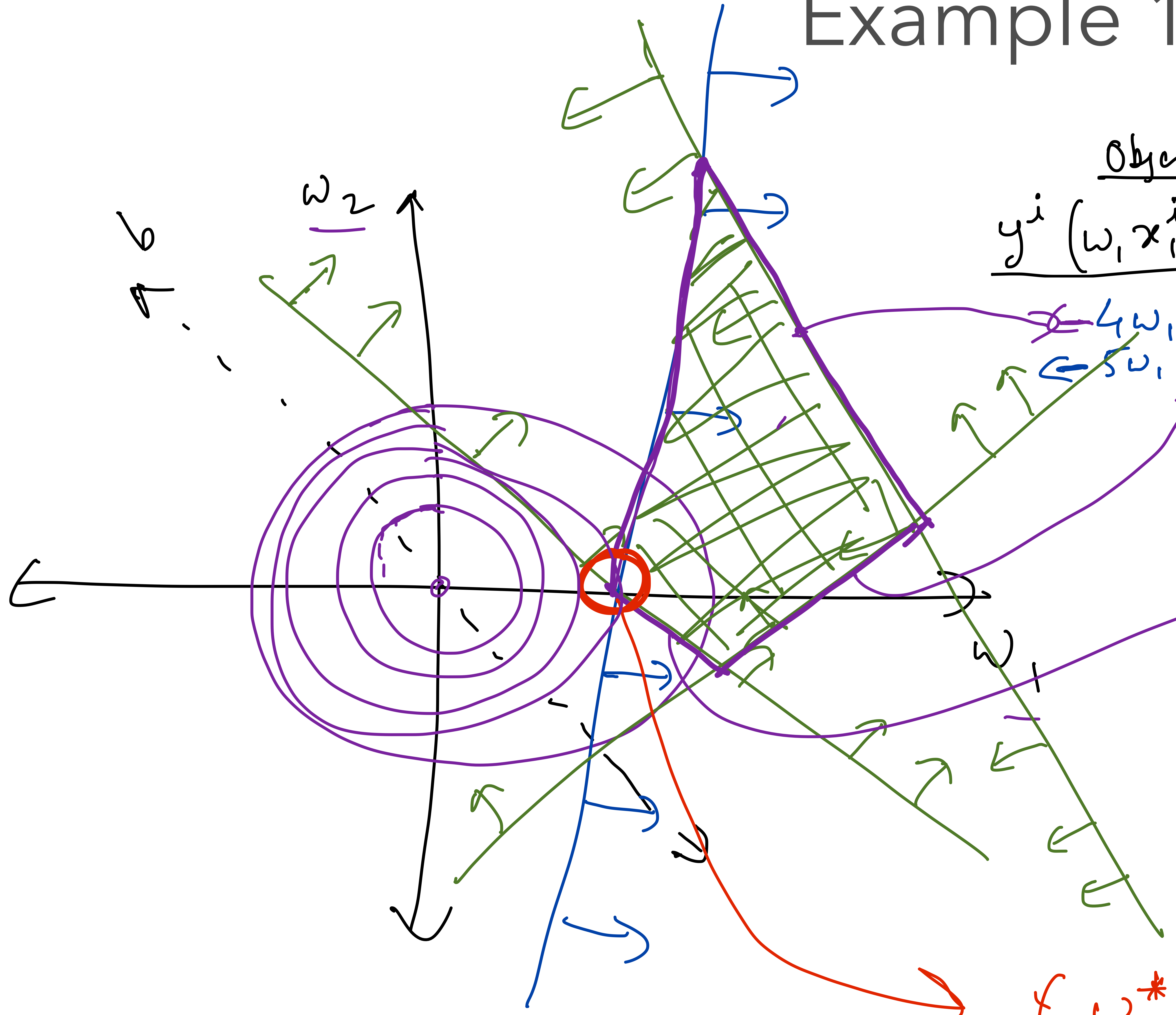
$4w_1 + 3w_2 + b \geq 1$

$5w_1 + 5w_2 + b \geq 1$

$2w_1 + w_2 + b \leq -1$

$w_1 + 3w_2 + b \leq -1$

Example 1



Objective: Optimize (w_1, w_2, b)

$y^i(w_1 x_1^i + w_2 x_2^i + b) \geq 1$ minimize $w_1^2 + w_2^2$

$4w_1 + 3w_2 + b \geq 1$

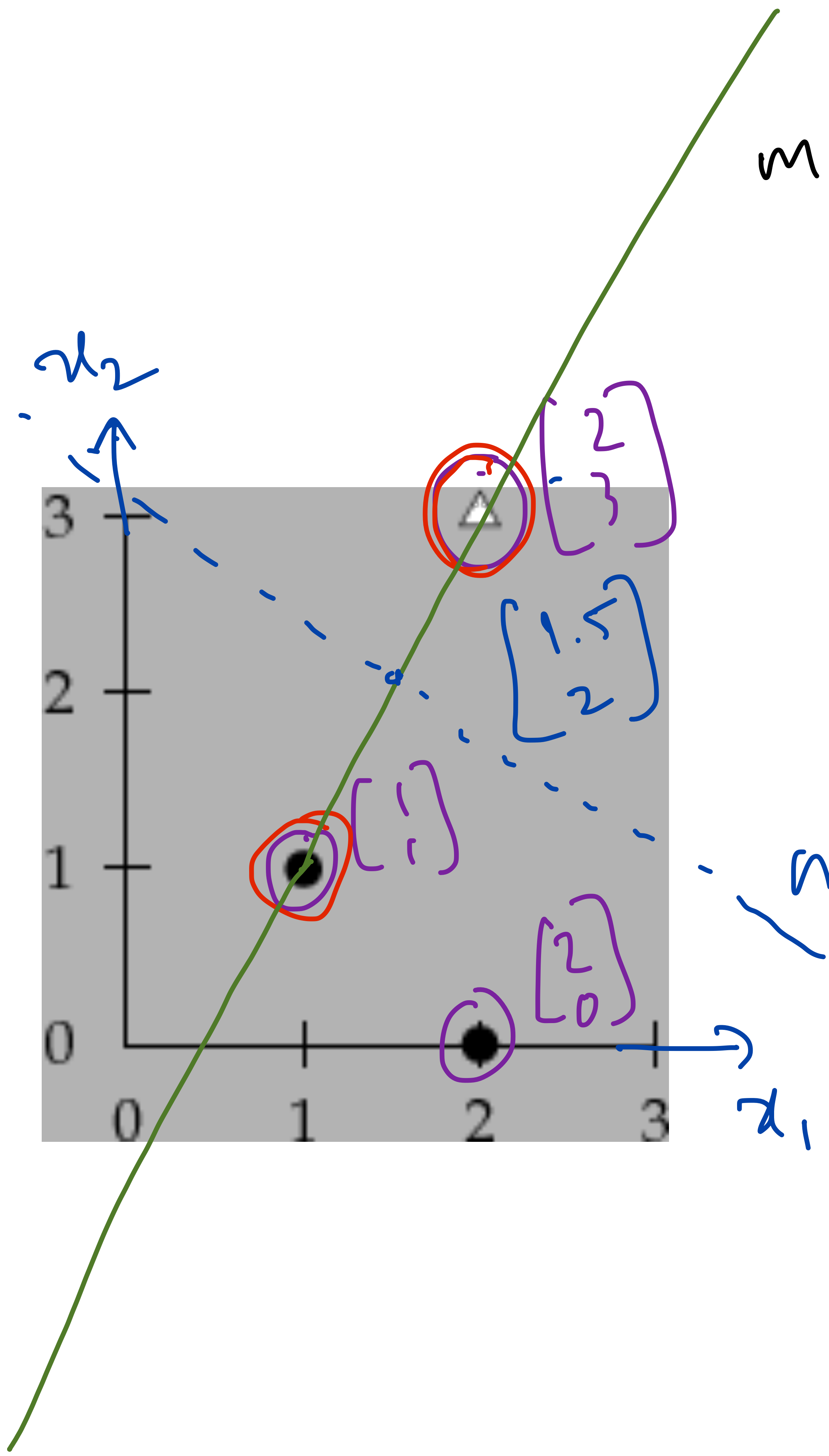
$5w_1 + 5w_2 + b \geq 1$

$-2w_1 + w_2 + b \leq -1$

$-w_1 + 3w_2 + b \leq -1$

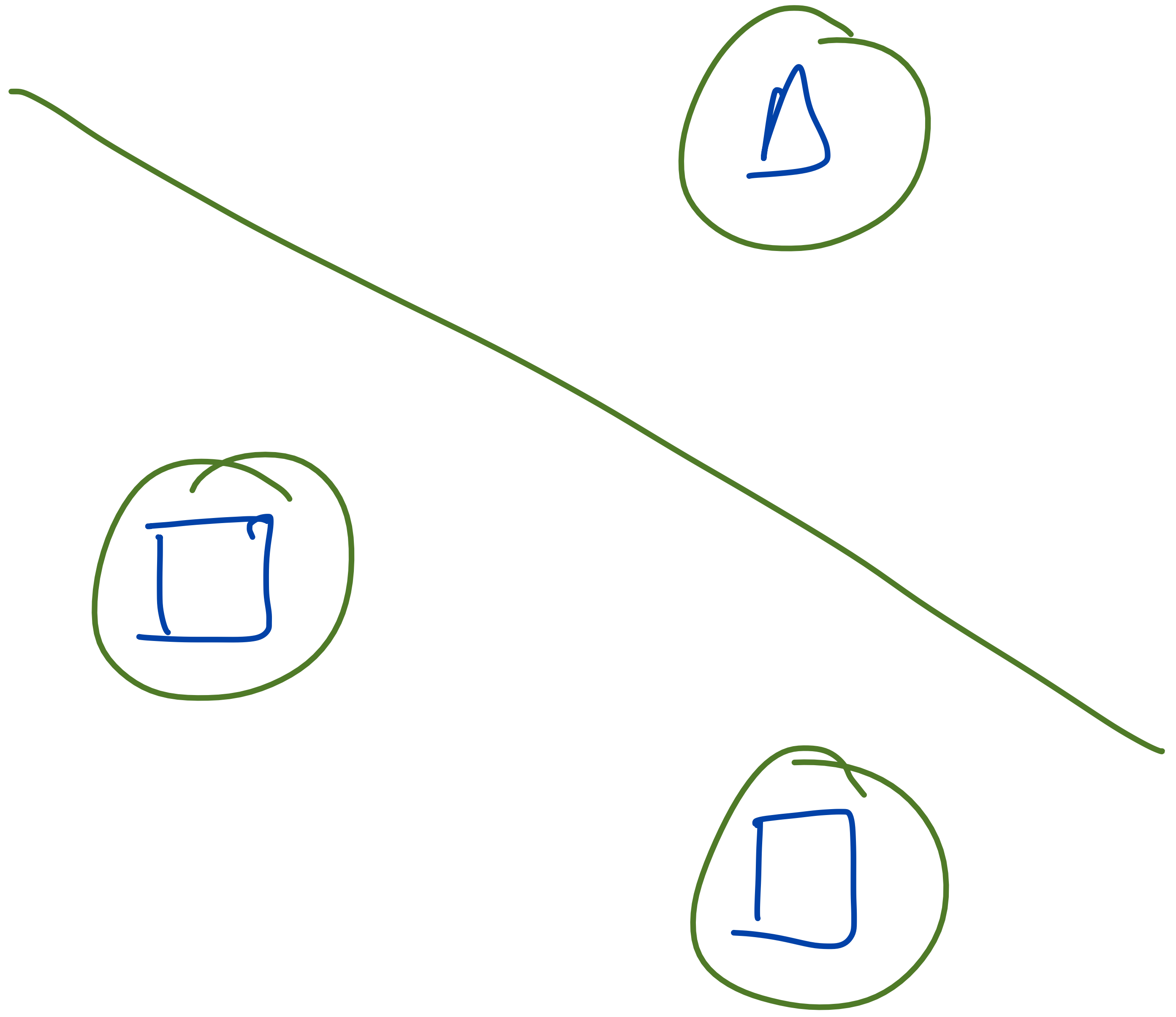
(w_1^*, w_2^*, b^*)

Example 2



$m = 2$

$m = -\frac{1}{2}$



$x_2 = -\frac{1}{2}x_1 + c$

$2x_2 + x_1 + b = 0 \Longrightarrow x_1 + 2x_2 - 5.5 = 0$

$4 + 1.5 + b = 0 \Rightarrow b = -5.5$

https://nlp.stanford.edu/IR-book/html/htmledition/support-vector-machines-the-linearly-separable-case-1.html

Online demo

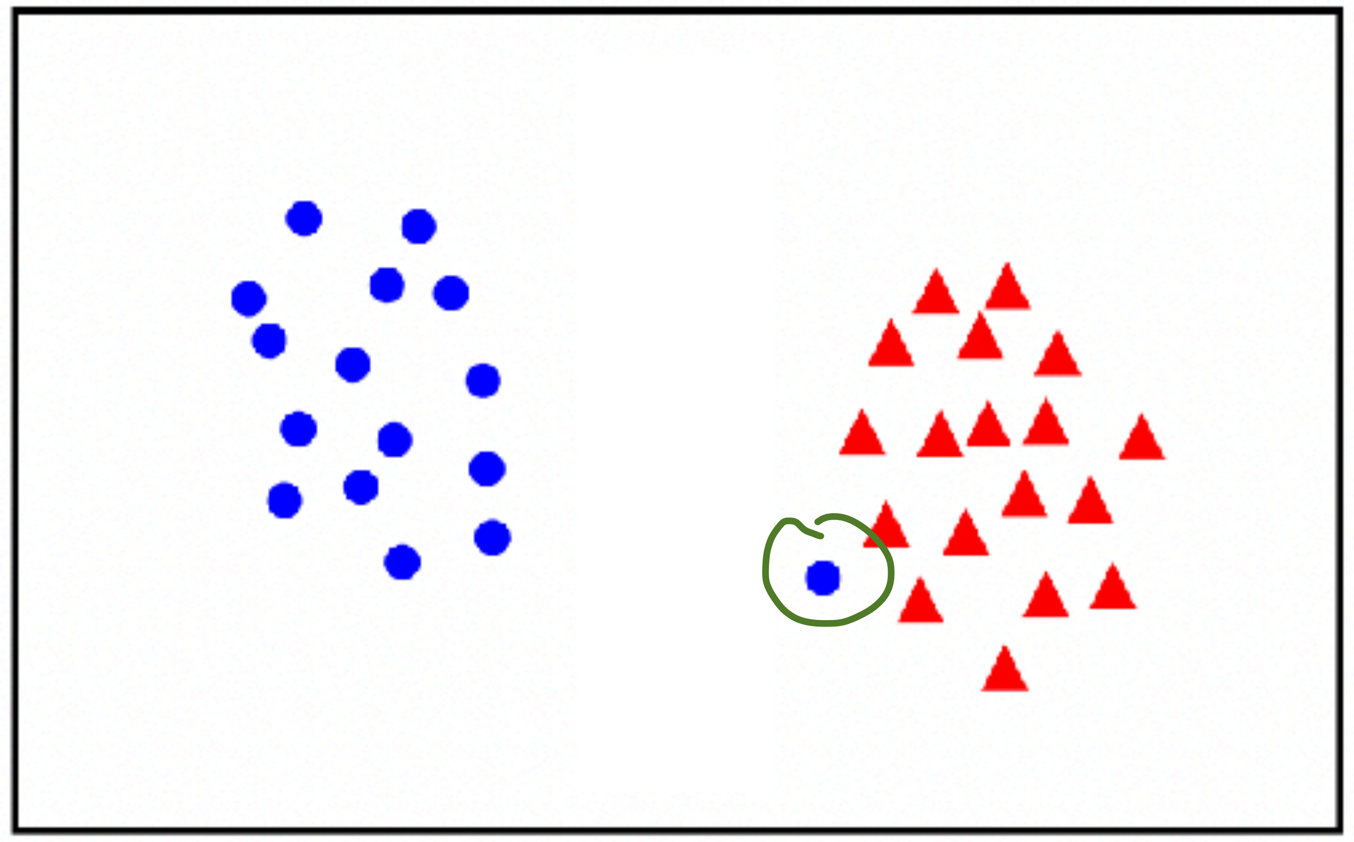
https://jgreitemann.github.io/svm-demo

How do we address data errors?

Credits: A. Zisserman (https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf)

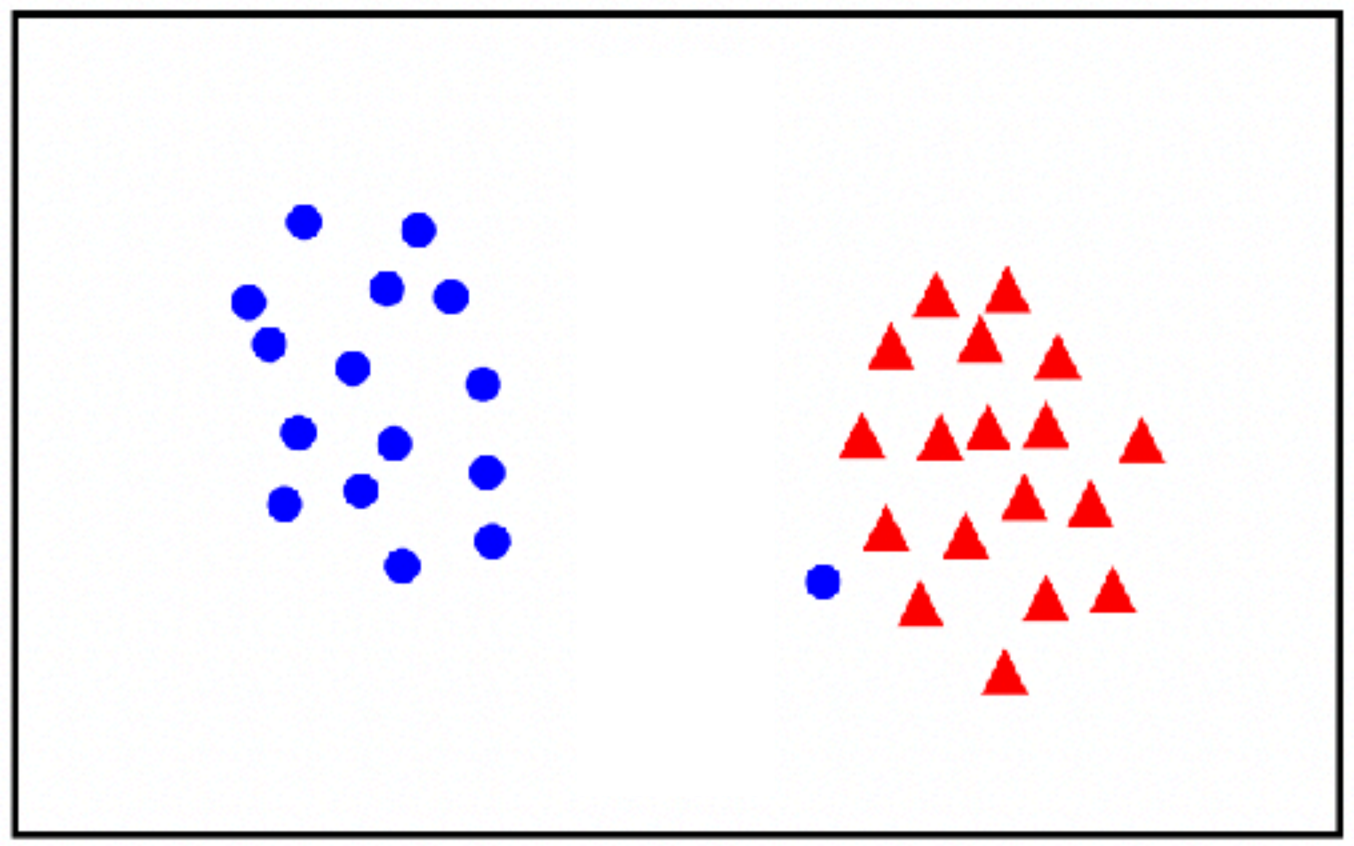
How do we address data errors?

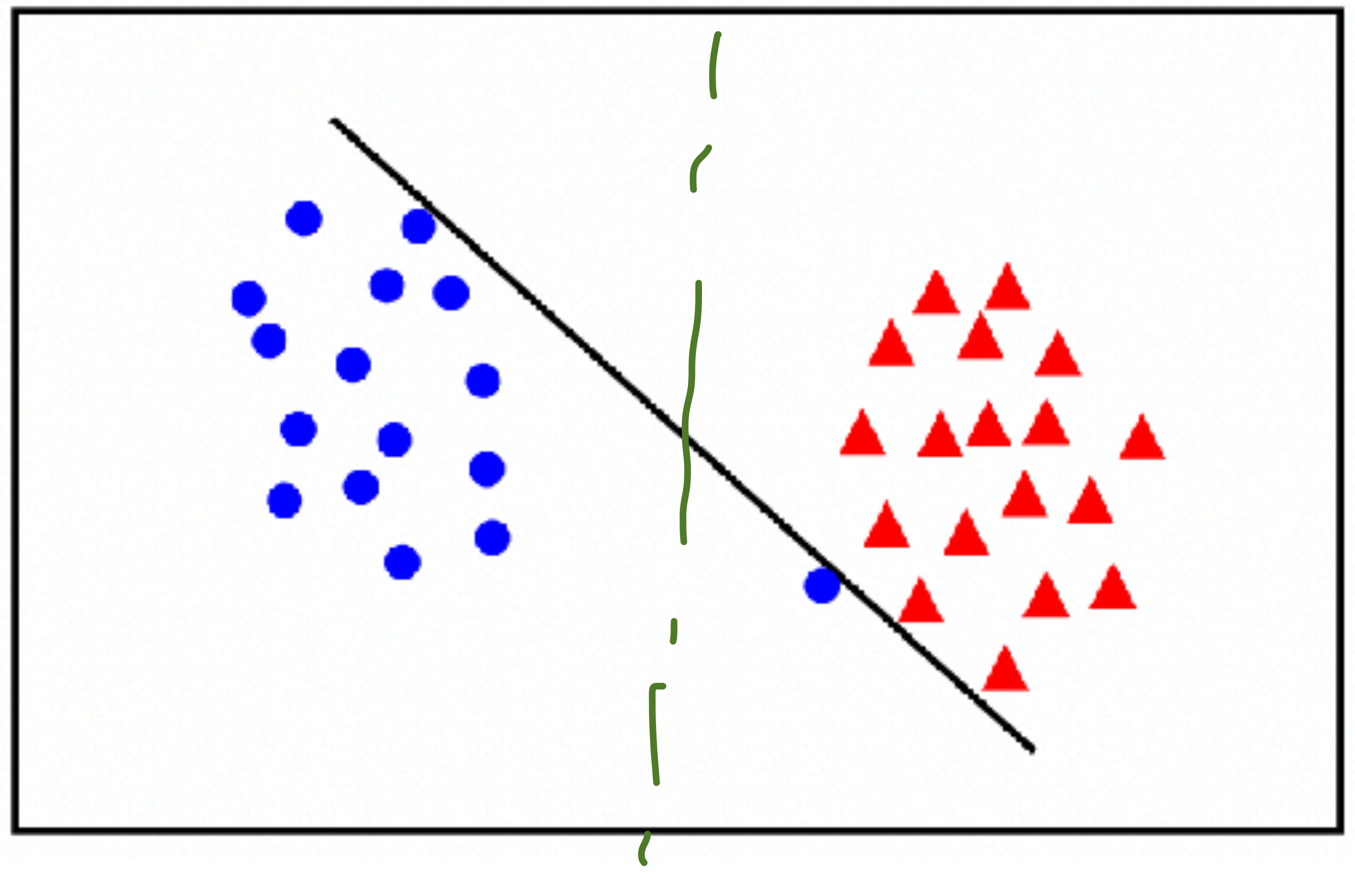
Dataset



Credits: A. Zisserman (https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf)

How do we address data errors?

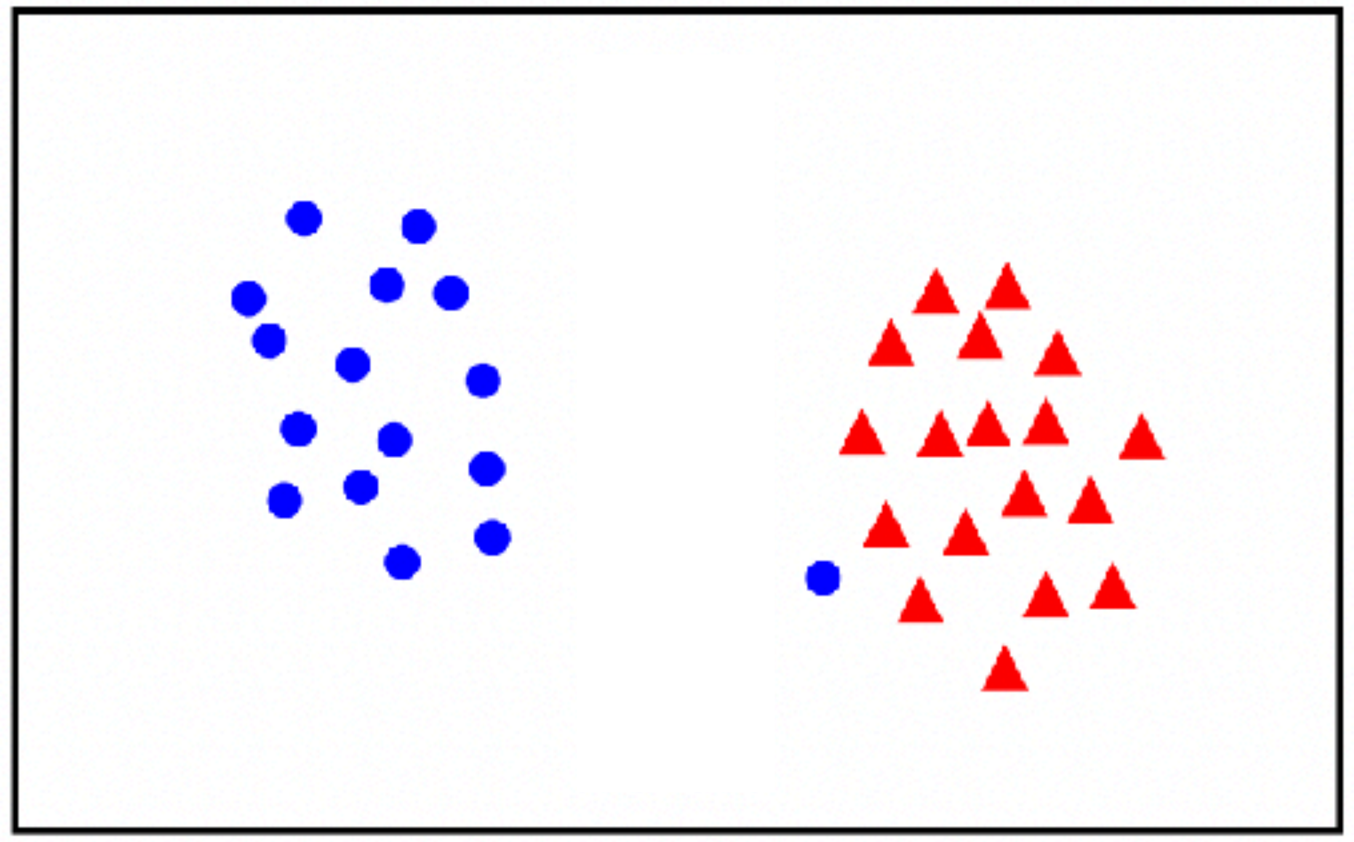




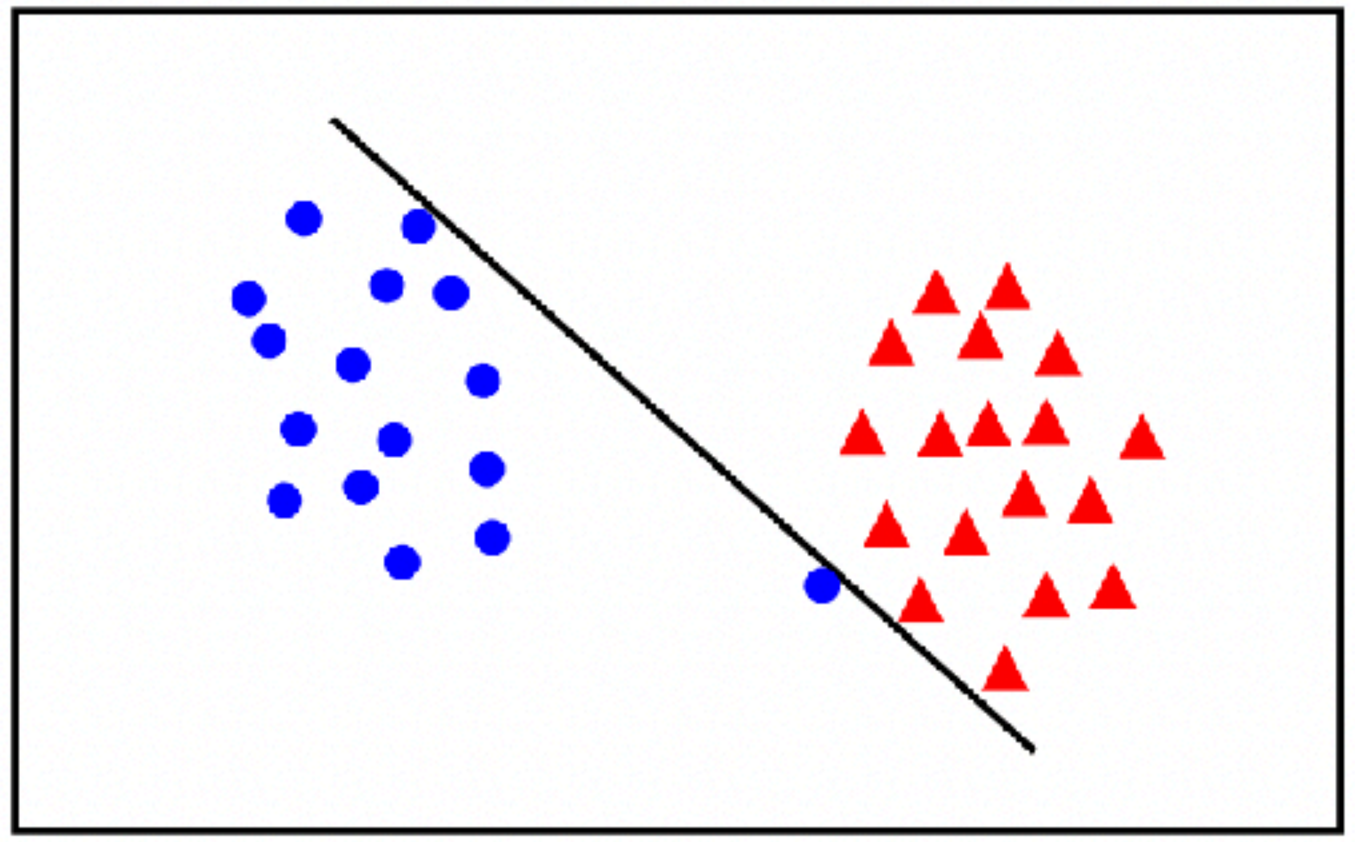
Credits: A. Zisserman (https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf)

How do we address data errors?

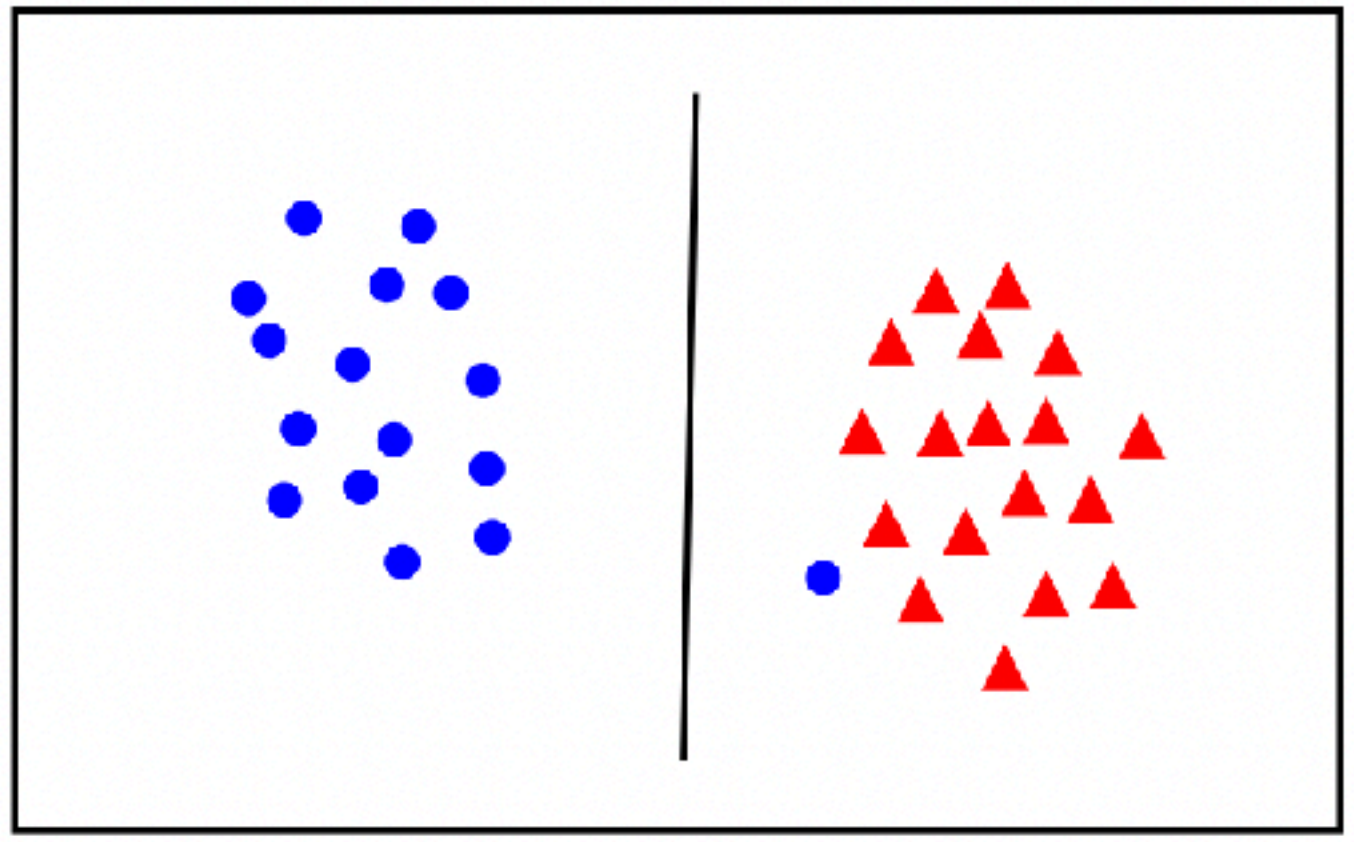
Dataset



Hard margin SVM



Soft margin SVM



Credits: A. Zisserman (https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf)

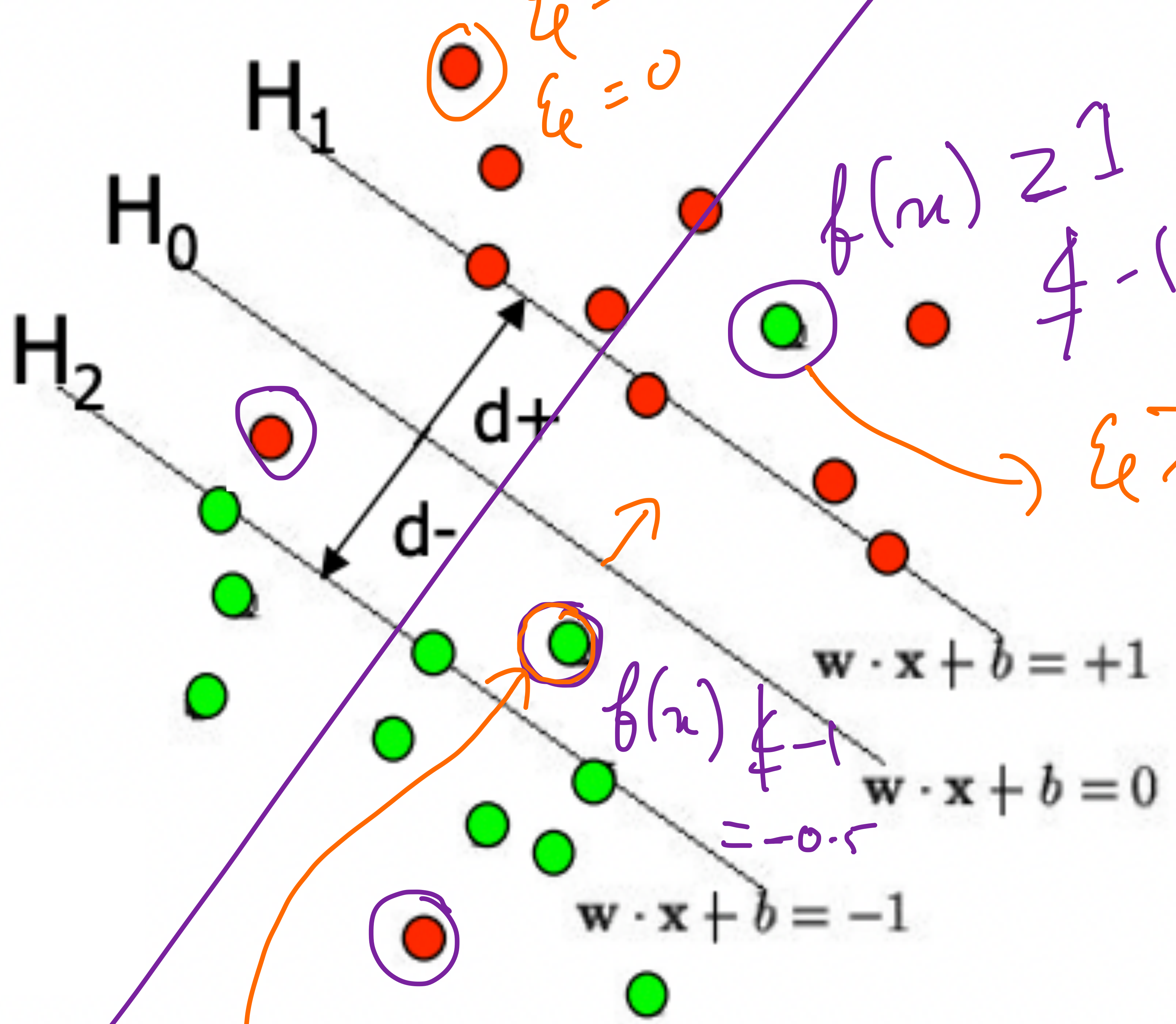
Handling margin violations

Credits: A. Zisserman (https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf)

$\boxed{y^i(w^Tx^i+b) \geq 1-\xi_i}$

Handling margin violations

$\xi_i = 0$



$\xi_i=0$ $\xi_i=0$

$f(x) \geq 1$ $\neq -1$

$\xi_i > 1$

$f(x) \neq -1$ $= -0.5$

$0 < \xi_i < 1$

- Goal:

max margin min slack

$$\min_{w\in\mathbb{R}^d,\xi_i\in\mathbb{R}} \|w\|^2 + C\sum_i^n \xi_i$$

$$\text{subject to } y^i(w^Tx^i+b) \geq 1-\xi_i$$

$$\text{and } \xi_i \geq 0$$

$C \to \infty$ Hard margin

$C \to 0$

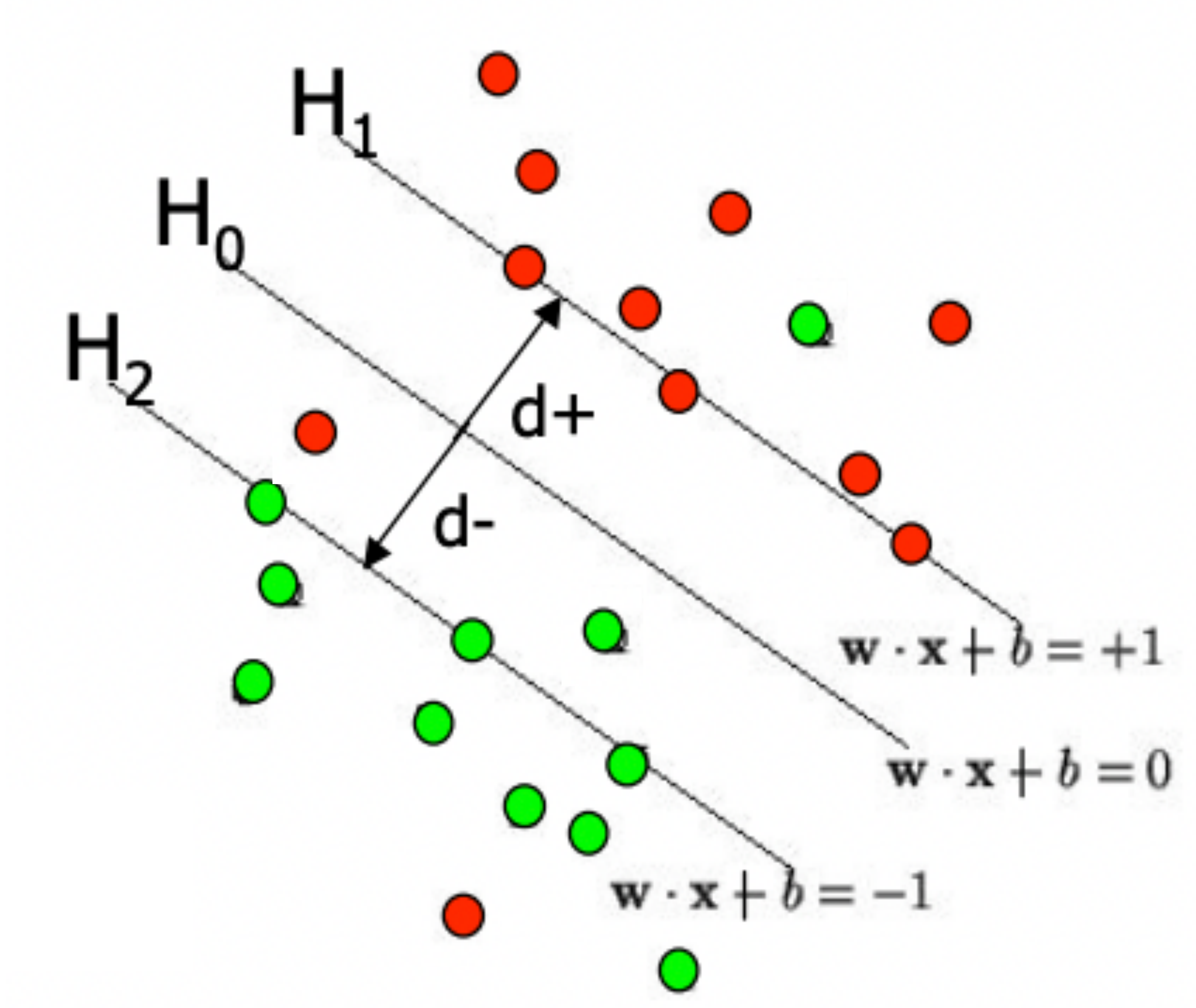
$\|w\|^2 = 0$

- Can be solved as a quadratic optimization problem with linear constraints.

w^*, b^*, ξ^*

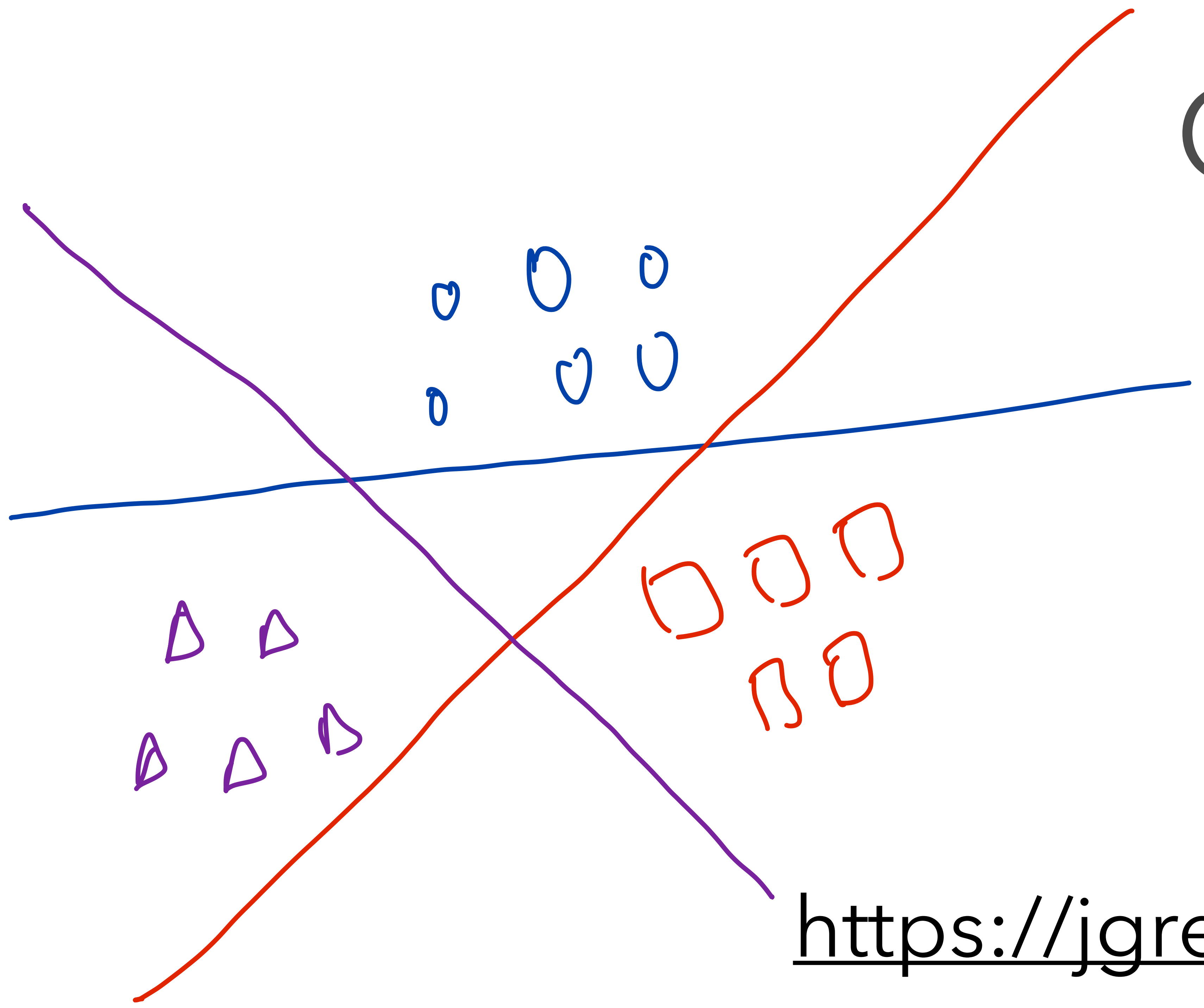
Credits: A. Zisserman (https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf)

Interpretation through Loss function



Credits: A. Zisserman (https://www.robots.ox.ac.uk/~az/lectures/ml/lect2.pdf)

Online demo



https://jgreitemann.github.io/svm-demo

Additional Reading

- Original paper: http://image.diku.dk/imagecanon/material/cortes_vapnik95.pdf
- http://pyml.sourceforge.net/doc/howto.pdf
- Quadratic Programming: https://scaron.info/blog/quadratic-programming-in-python.html

Questions?