Introduction to Machine Learning
[Fall 2022]

Support Vector Machines (Part 3)

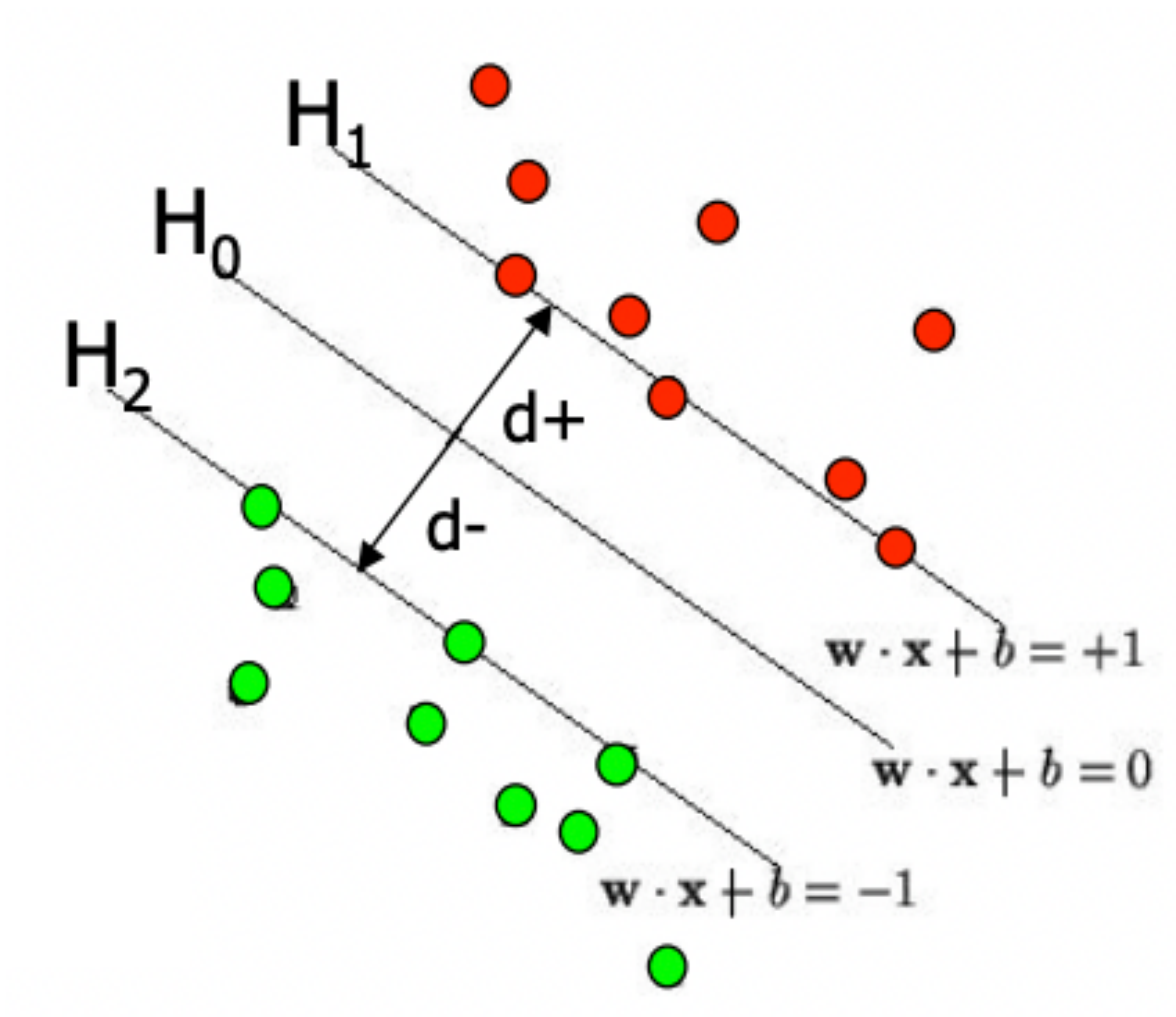
September 29, 2022

Lerrel Pinto

Topics for today

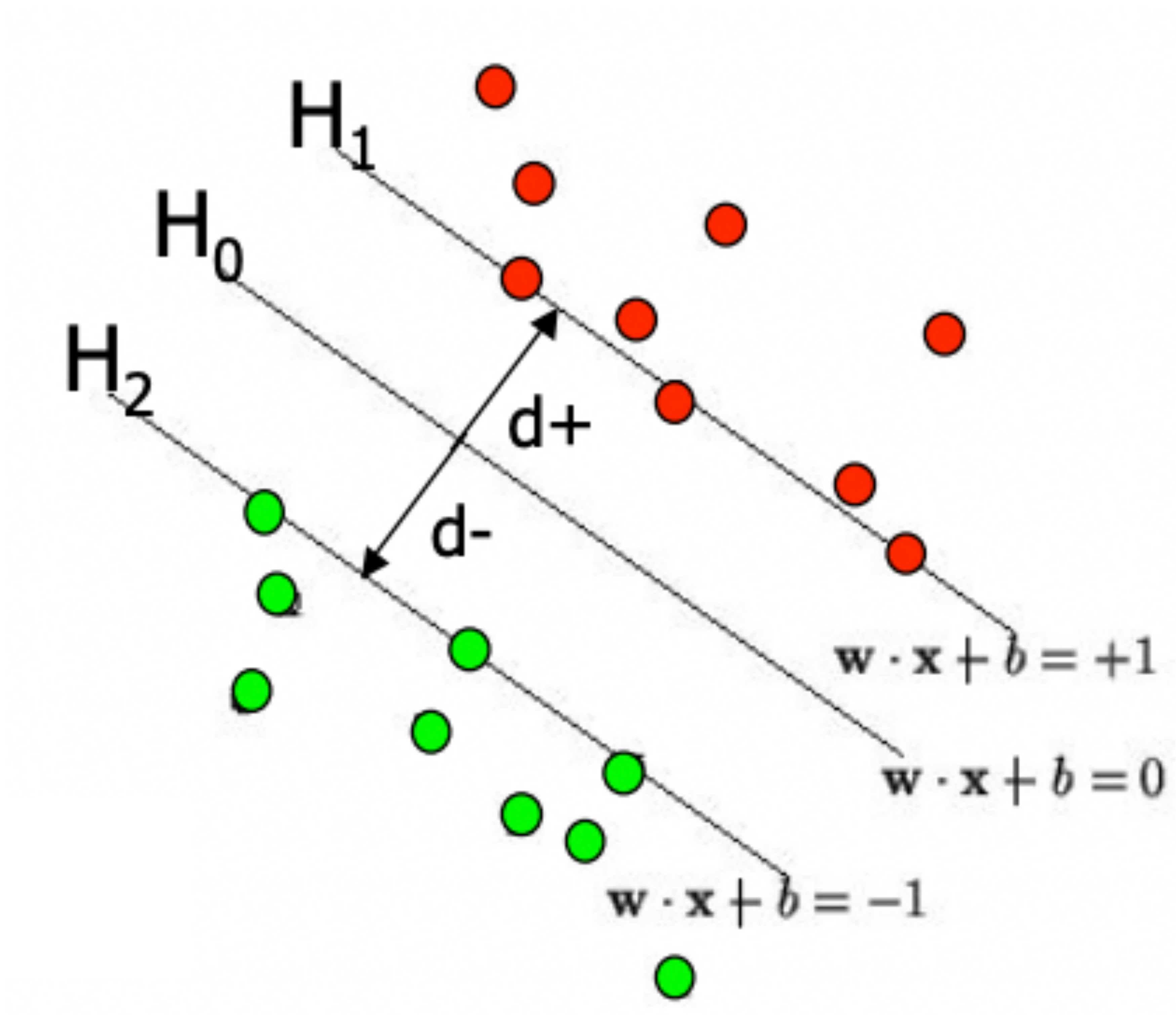
- Diving a little deeper into solving SVMs
- Kernel SVMs
- Evaluating a classifier

Recap: SVMs



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).

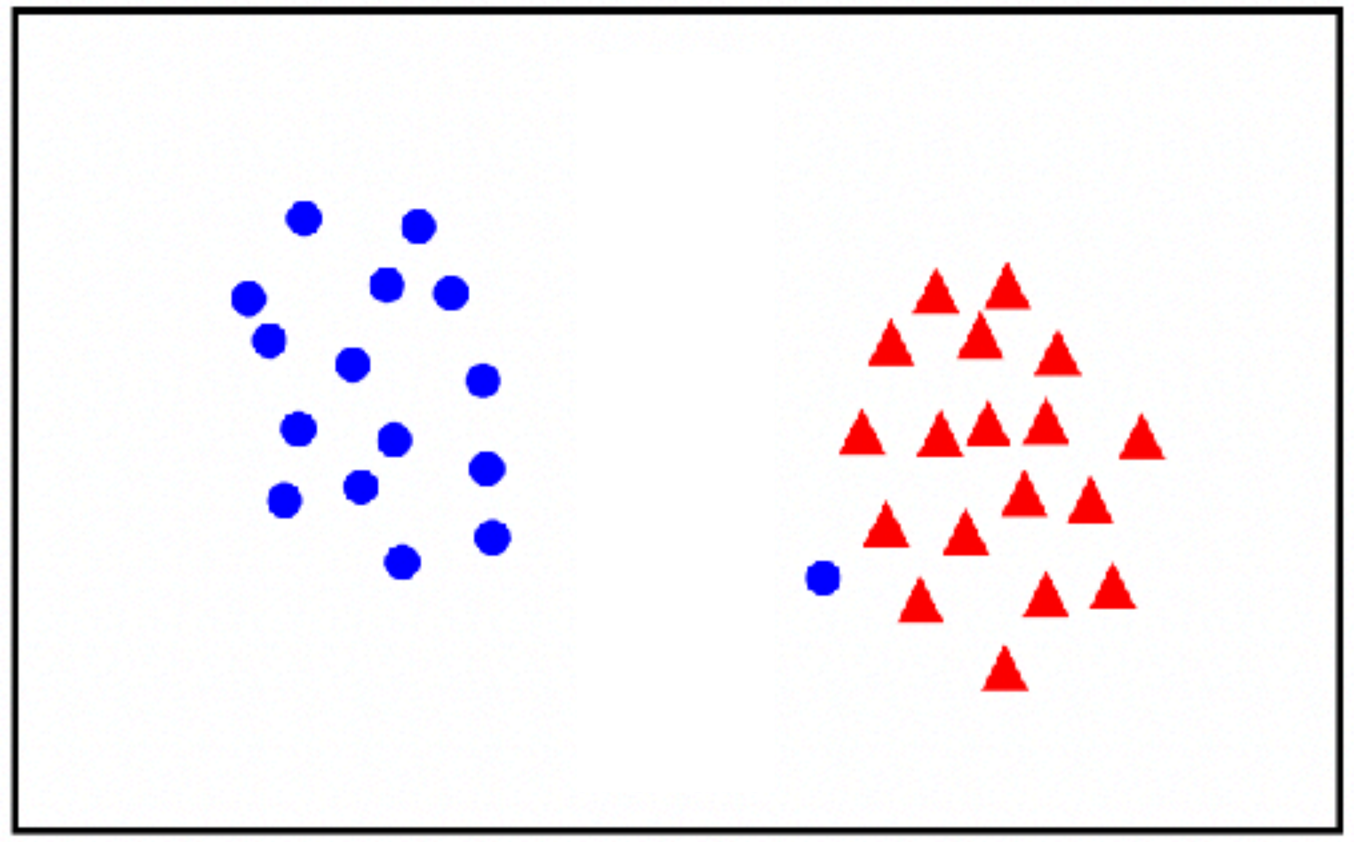
$$\boxed{\min_{w} ||w||^2 \text{ subject to } y^i(w^T x^i + b) \geq 1}$$

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

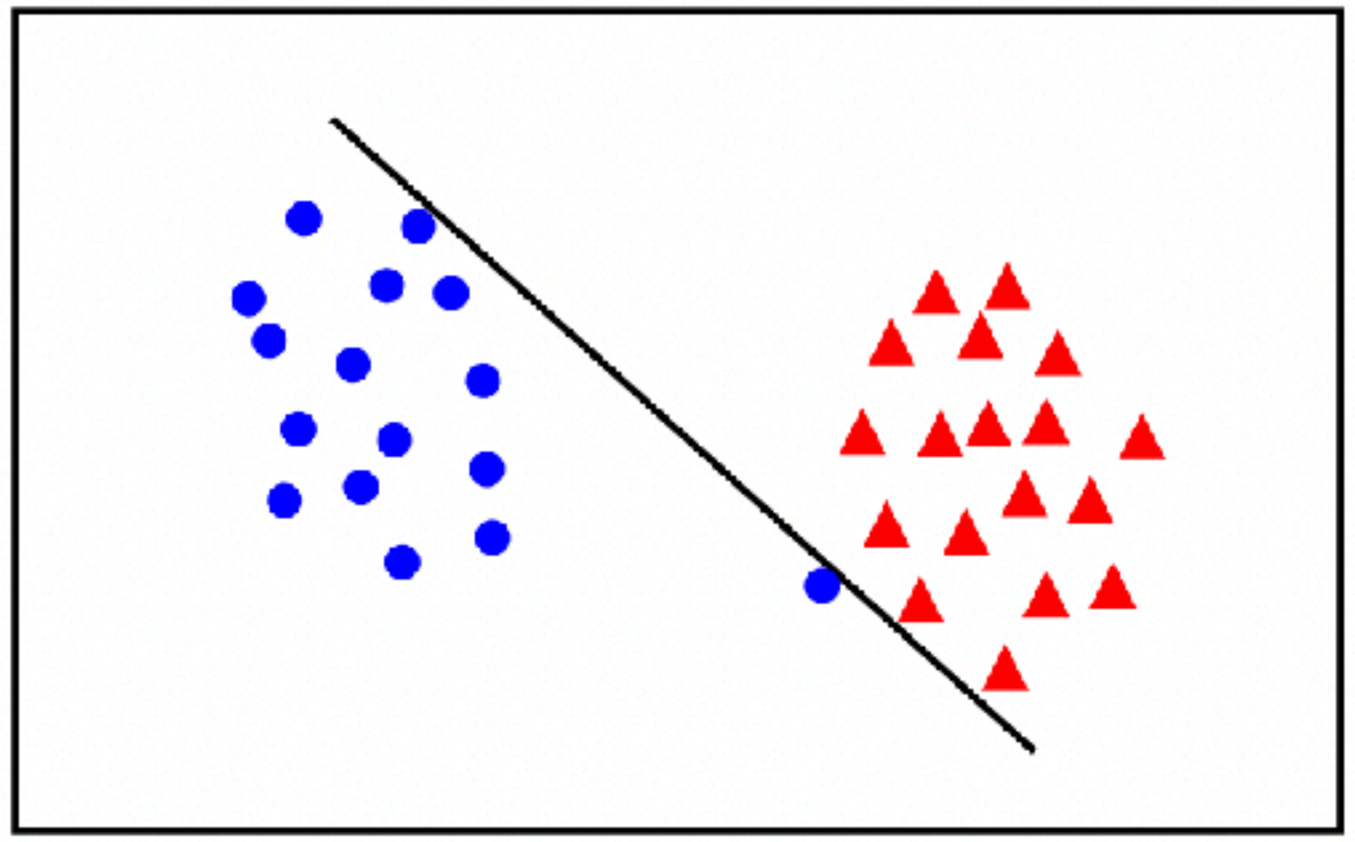
Credits: R. Berwick (https://web.mit.edu/6.034/wwwbob/svm-notes-long-08.pdf)

Recap: How do we address data errors?

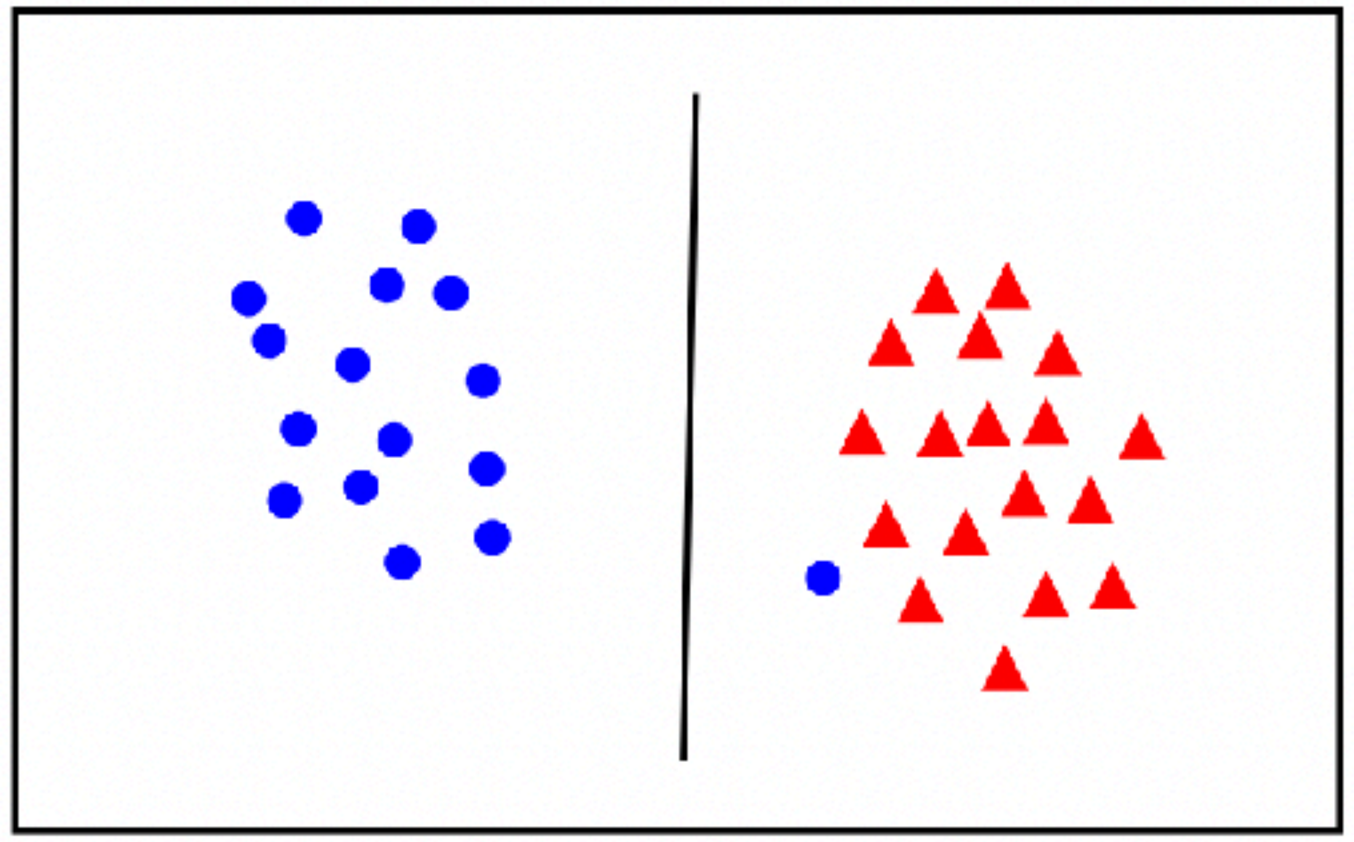
Dataset



Hard margin SVM

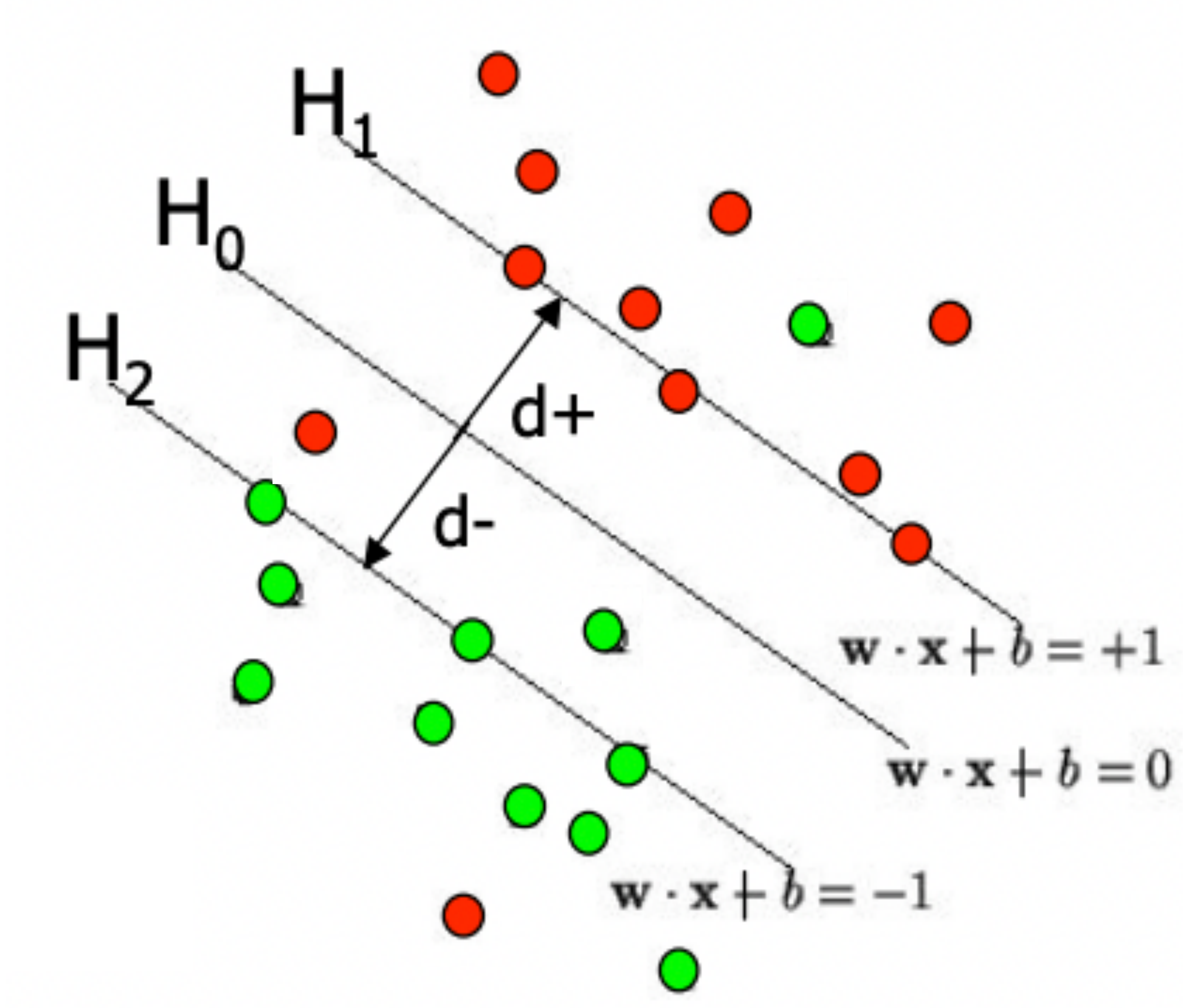


Soft margin SVM



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

Recap: Handling margin violations



- Goal:

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

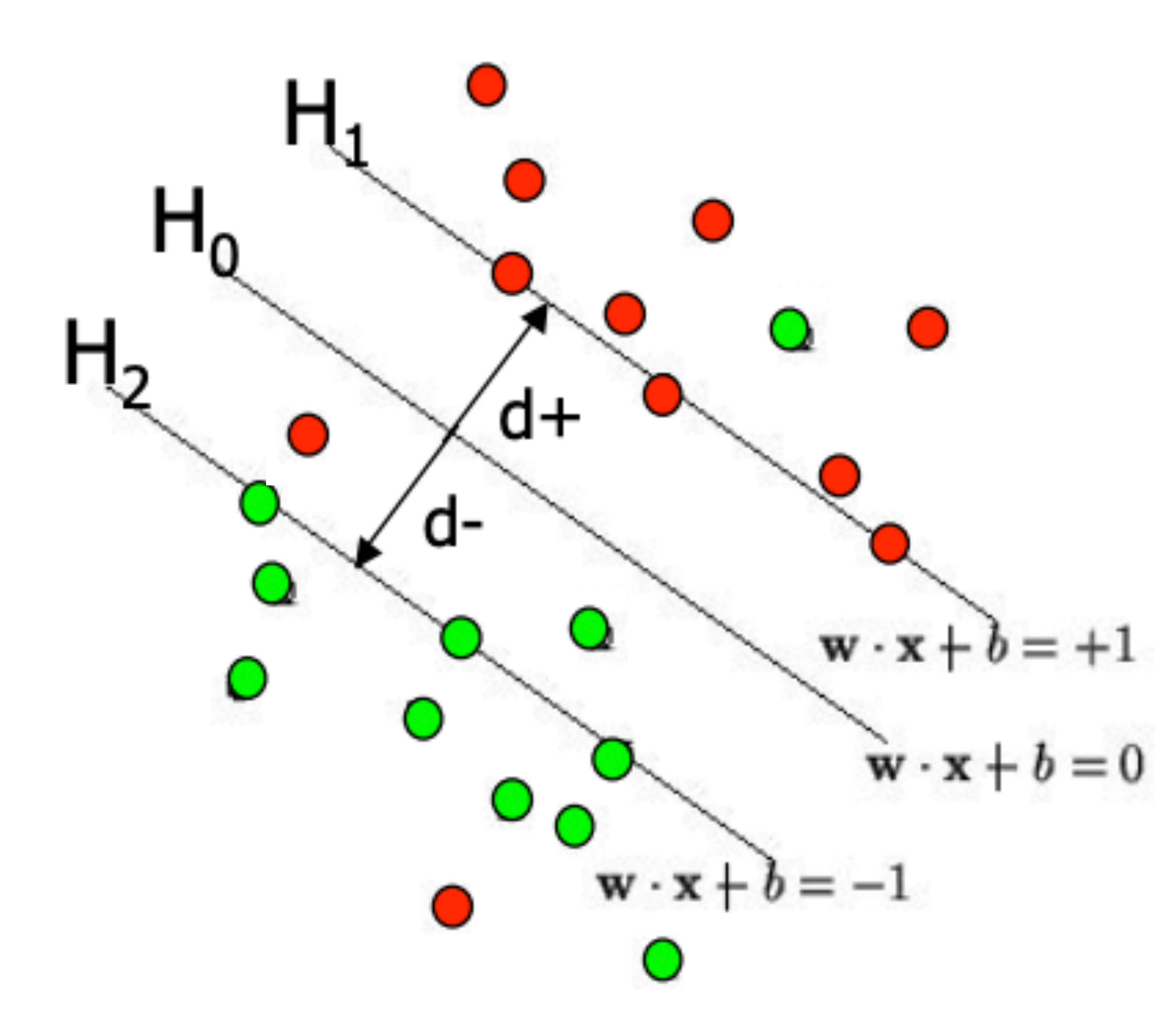
subject to $y^i(w^T x^i + b) \geq 1 - \xi_i$

and $\xi_i \geq 0$

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

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)

What does the SVM math look like?

Reframing the objective

Reframing the objective

Original Goal: Maximize margin / Minimize $||w||^2$

- Also need to satisfy $y^i f(x^i) \geq 1$ for all datapoints (x^i, y^i).

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

Reframing the objective

Original Goal: Maximize margin / Minimize $||w||^2$

- Also need to satisfy $y^i f(x^i) \geq 1$ for all datapoints (x^i, y^i).

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

Combined objective (aka Lagrangian):

$$L(w, \alpha) = ||w||^2 - \sum_i \alpha_i [y^i (w^T x^i + b) - 1];\ \ \alpha_i \geq 0, \forall i$$

Reframing the objective

Original Goal: Maximize margin / Minimize $||w||^2$

- Also need to satisfy $y^i f(x^i) \geq 1$ for all datapoints (x^i, y^i).

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

Combined objective (aka Lagrangian):

$$L(w, \alpha) = ||w||^2 - \sum_i \alpha_i [y^i(w^T x^i + b) - 1]; \;\; \alpha_i \geq 0, \forall i$$

New goal: $\min_{w,b} \max_{\alpha_i \forall i} L(w, a)$

Primal and Dual

$$L(w, \alpha) = \|w\|^2 - \sum_i \alpha_i [y^i (w^T x^i + b) - 1]; \;\; \alpha_i \geq 0, \forall i$$

New goal: $\min_{w,b} \max_{\alpha_i \forall i} L(w, a)$ – Primal

Primal and Dual

$$L(w, \alpha) = \|w\|^2 - \sum_i \alpha_i [y^i (w^T x^i + b) - 1]; \;\; \alpha_i \geq 0, \forall i$$

New goal: $\min\limits_{w,b} \max\limits_{\alpha_i \forall i} L(w, a)$ – Primal

Equivalent goal: $\max\limits_{\alpha_i \forall i} \min\limits_{w,b} L(w, a)$ – Dual

Primal and Dual

$$L(w,\alpha) = \|w\|^2 - \sum_i \alpha_i[y^i(w^T x^i + b) - 1];\ \ \alpha_i \geq 0, \forall i$$

New goal: $\min_{w,b} \max_{\alpha_i \forall i} L(w,a)$ – Primal

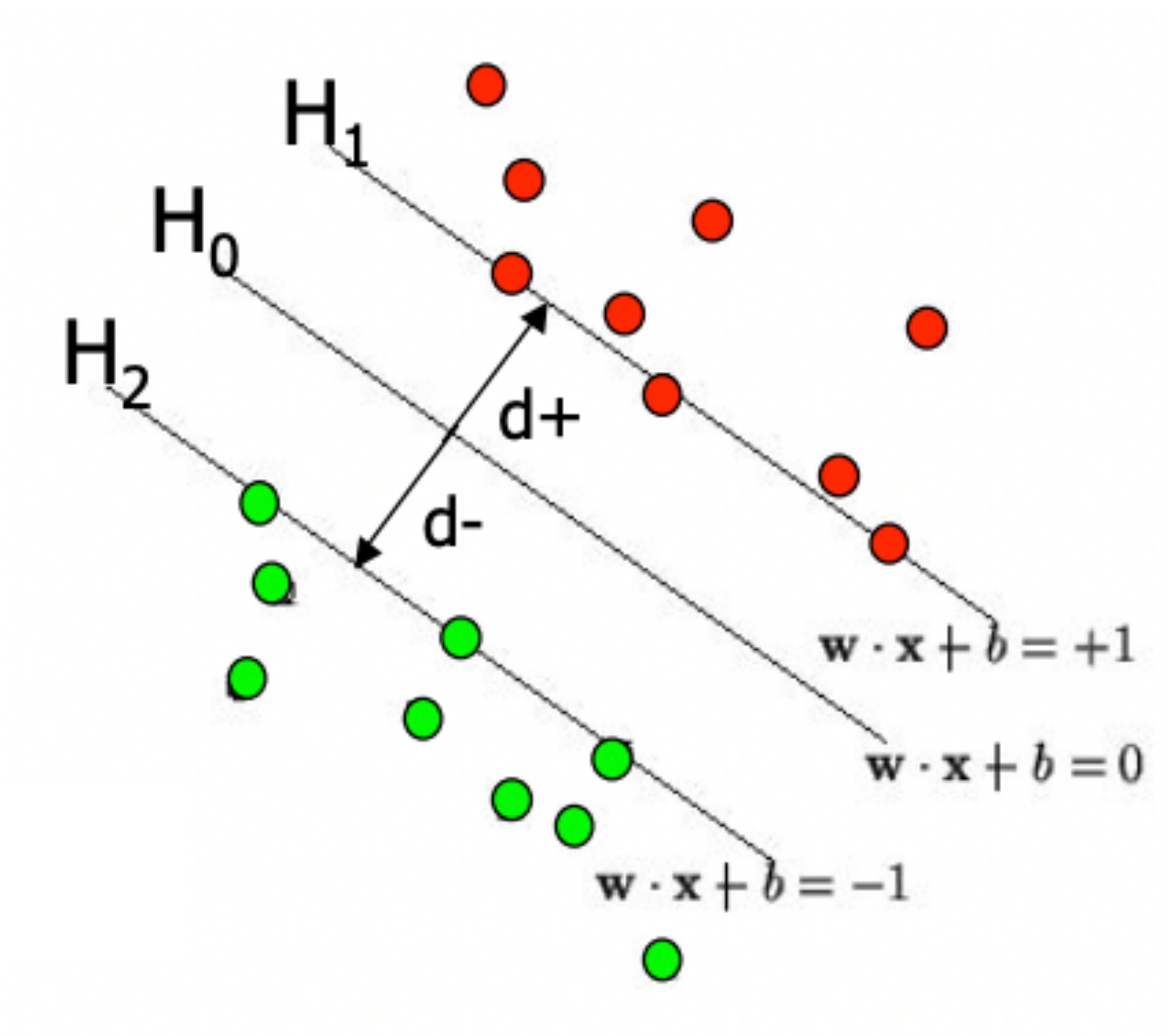
Equivalent goal: $\max_{\alpha_i \forall i} \min_{w,b} L(w,a)$ – Dual

Solution from setting derivatives to zero:

- $w = \sum_i \alpha_i y_i x_i$ and $\sum_i \alpha_i y_i = 0$ – one step of math

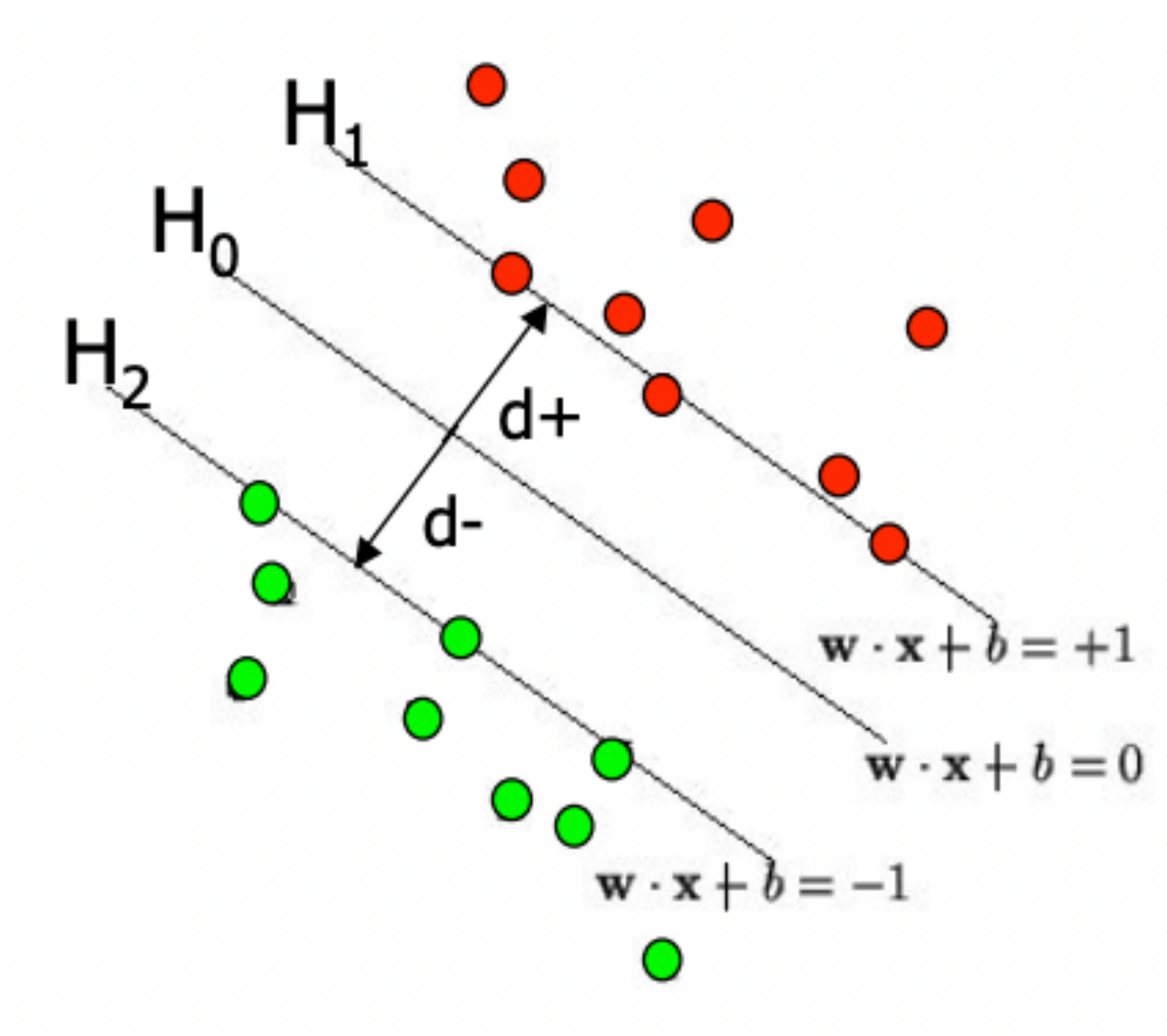
  $b = y^k - w^T x^k$ For any k where $\alpha_k > 0$– several steps of math

Looking back SVMs



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

Looking back SVMs



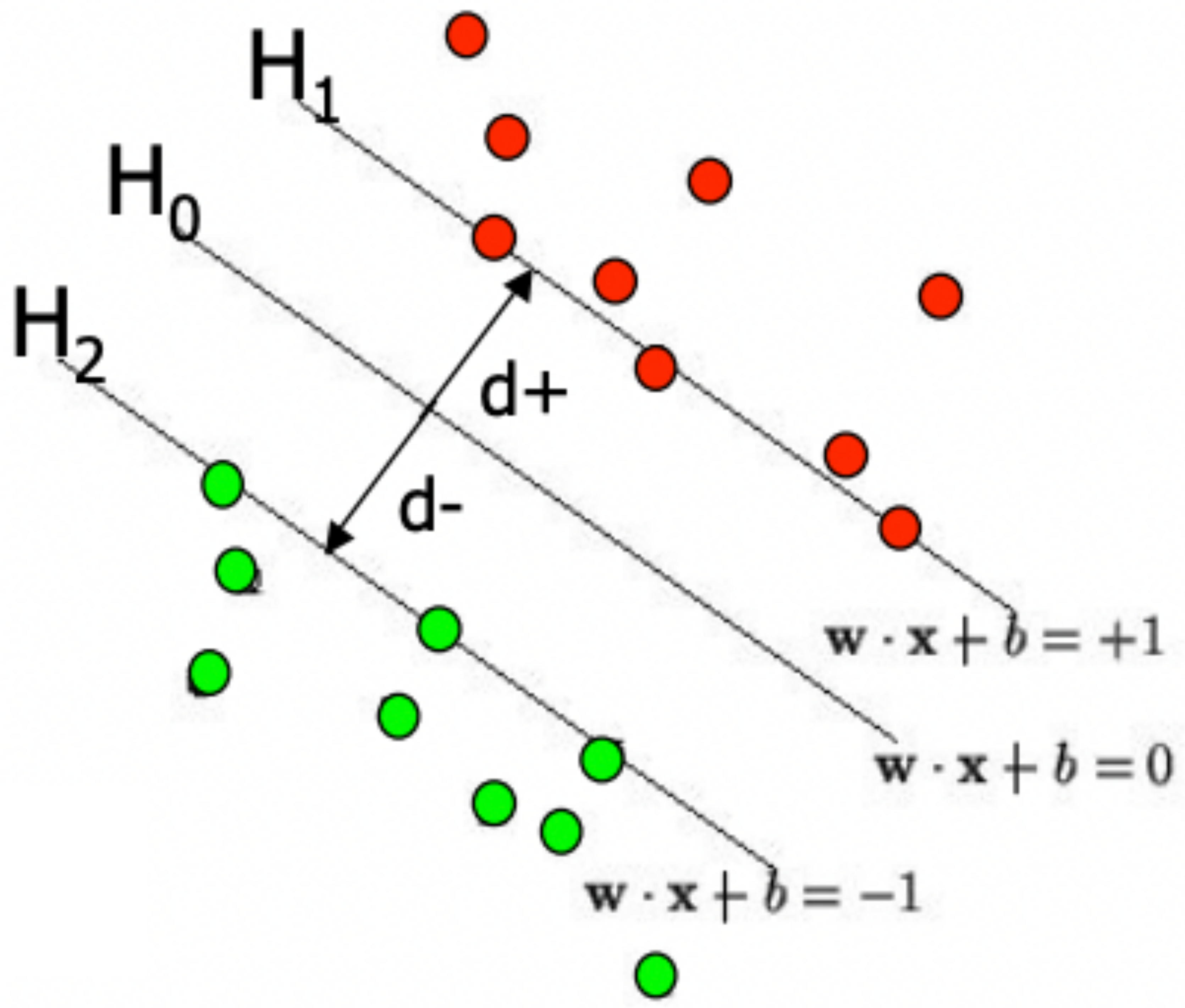
$$w = \sum_i \alpha_i y_i x_i$$

$$b = y^k - w^T x^k$$

For any k where $\alpha_k > 0$

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

Looking back SVMs



$$w = \sum_i \alpha_i y_i x_i$$

$$b = y^k - w^T x^k$$

For any k where $\alpha_k > 0$

$$f(x) = \sum_i \alpha_i y^i (x^i . x) + b$$

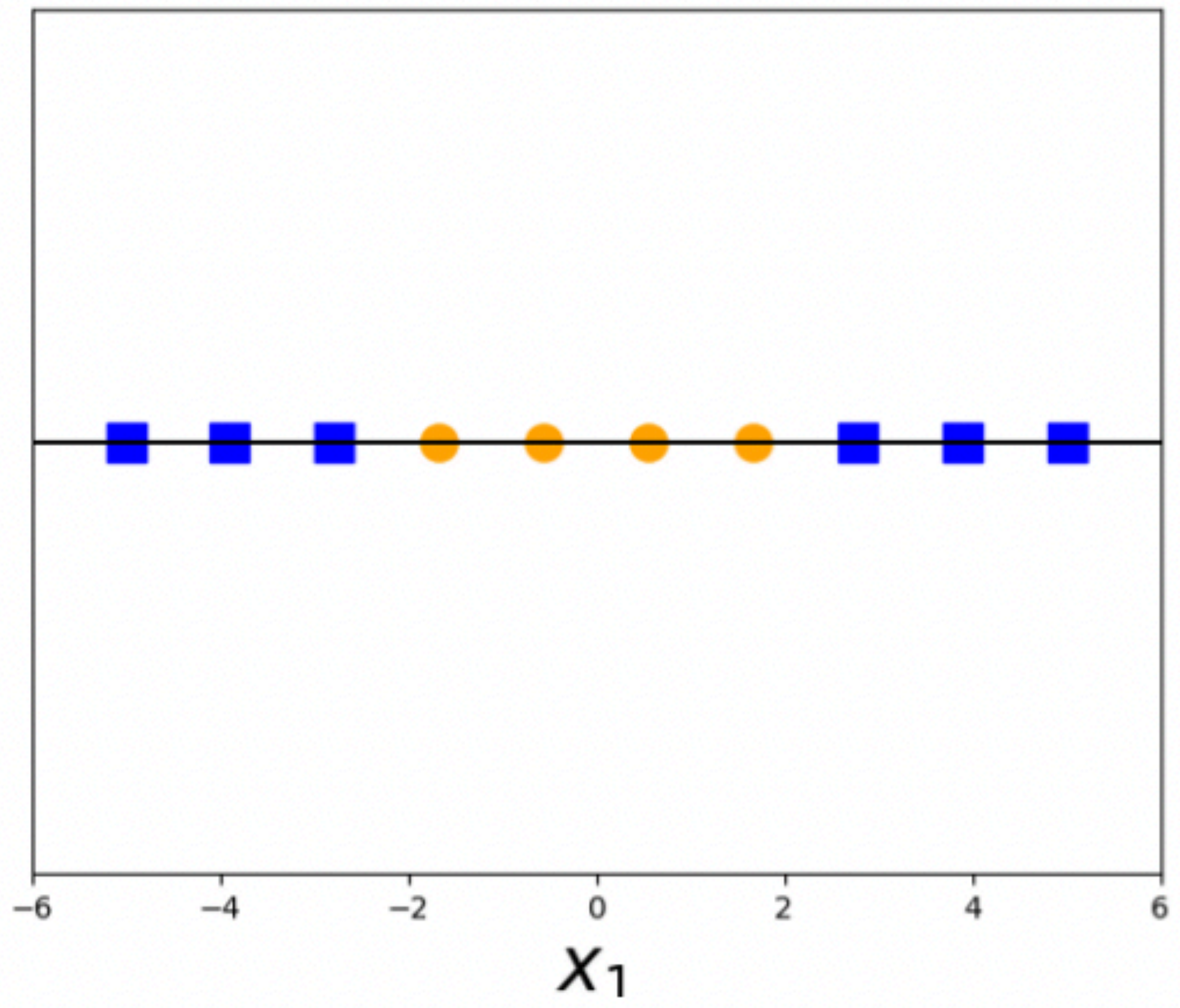
$$\hat{y} = sign(f(x))$$

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

How to handle problems where the data is non-linear?

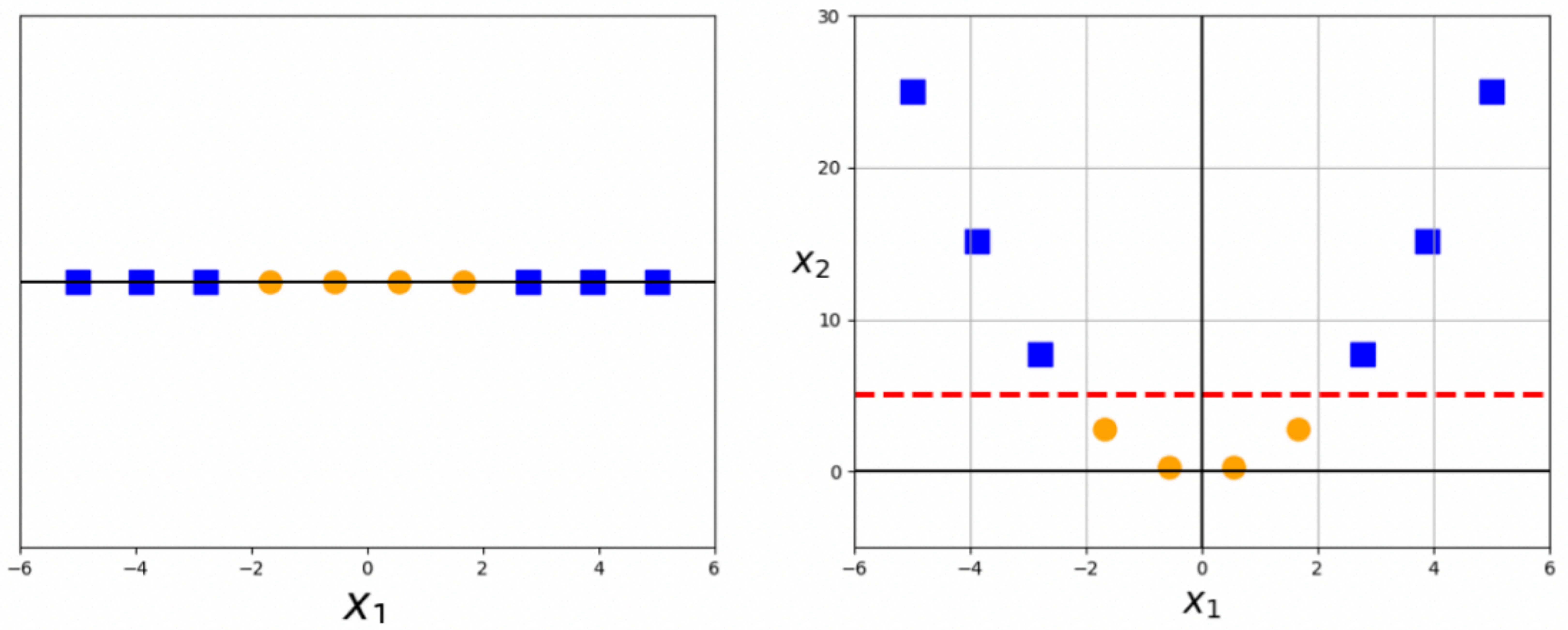
Credits: https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f

How to handle problems where the data is non-linear?



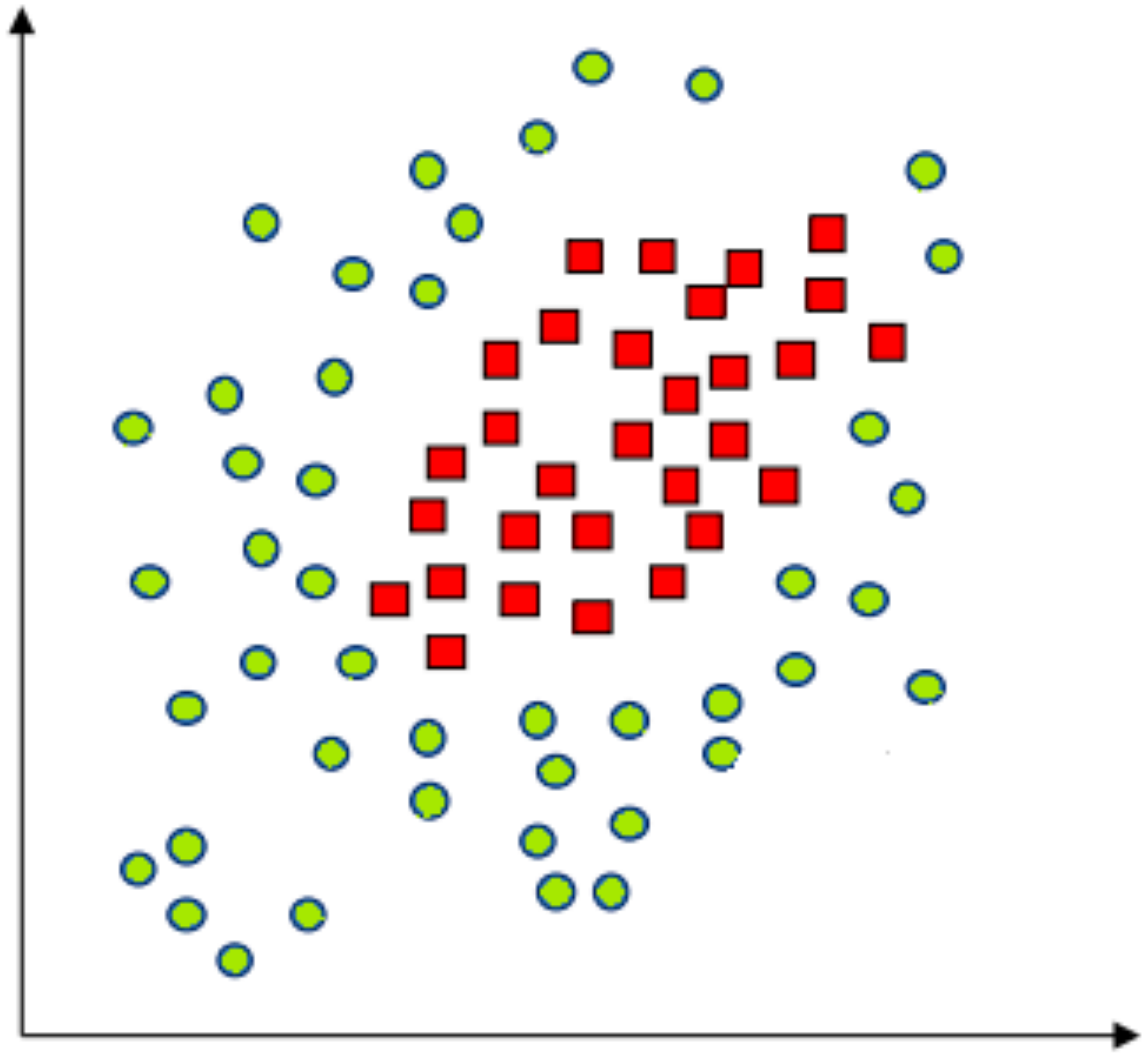
Credits: https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f

How to handle problems where the data is non-linear?



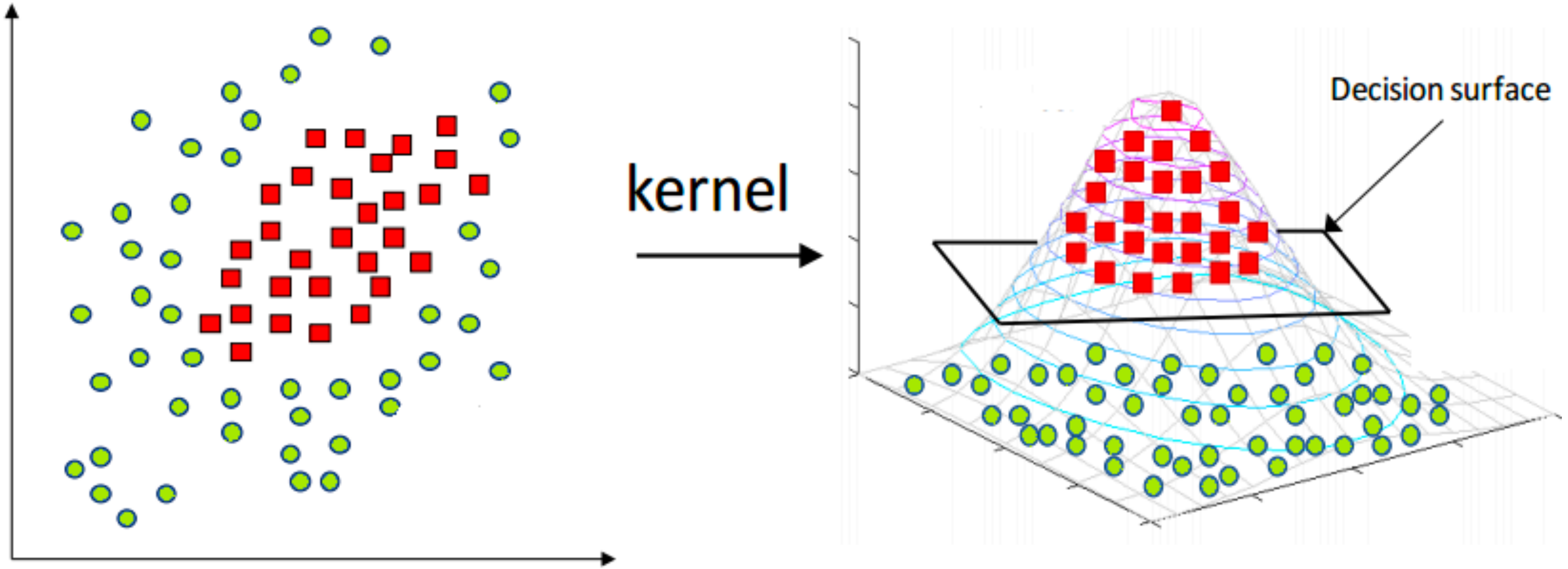
Credits: https://towardsdatascience.com/the-kernel-trick-c98cdbcaeb3f

How to handle problems where the data is non-linear?



Credits: https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d

How to handle problems where the data is non-linear?



Credits: https://medium.com/@zxr.nju/what-is-the-kernel-trick-why-is-it-important-98a98db0961d

Cool math trick – Kernels

$$\hat{y} = sign(\sum_i \alpha_i y^i (x^i \,.\, x) + b)$$

Linear case

$$\hat{y} = sign(\sum_i \alpha_i y^i (\phi(x)^i \,.\, \phi(x)) + b)$$

Non-linear case

$$\hat{y} = sign(\sum_i \alpha_i y^i (K(x^i, x)) + b)$$

Non-linear with kernels

Cool math trick – Kernels

$$k(x, x') = x^T x'$$

Linear kernel

$$k(x, x') = (1 + x^T x')^d$$

Polynomial kernel

$$k(x, x') = exp(-\|x - x'\|^2 / 2\sigma^2)$$

Gaussian kernel

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
- Lecture notes: https://www.robots.ox.ac.uk/~az/lectures/ml/lect3.pdf
- Lecture notes: http://people.csail.mit.edu/dsontag/courses/ml13/slides/lecture6.pdf

Questions?