CIS 419/519 Recitation

Chang Liu

Nov 11 2020

Content

- VC Dimension
- SVM
- Adaboost

Part 1: VC Dimension

Definitions

- We say that a set S of examples is shattered by a set of functions H if for every partition of the examples in S into positive and negative examples there is a function in H that gives exactly these labels to the examples

- The VC dimension of hypothesis space H over instance space X is the size of the largest finite subset of X that is shattered by H.

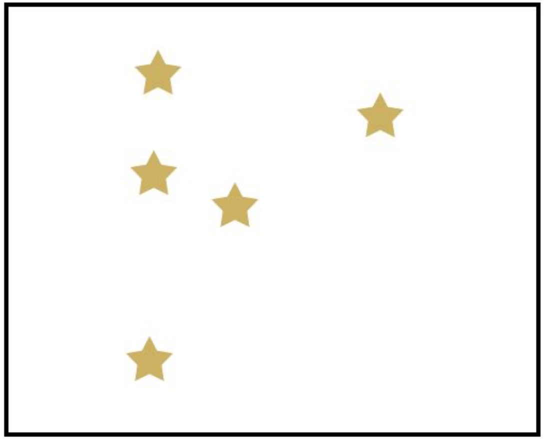
Shatter?

Size of largest finite subset of X?

Walkthrough Example

Setting:

- We have 5 data points scatter randomly in a 2D space
- We propose a linear separator, in Hypothesis Space (H)
- Objective: Find VC(H)



Strategy Summary:

We will use a different strategy:

1. Guess the VC dimension (in this case, we guess 3)
2. Find a set of size 3 that is shattered by H
3. Show that no set of size 4 is shattered by H

This is enough to show that the biggest set shattered by H has size 3

Why we want it?

- Help us to figure out how expressive a Hypothesis Space is, especially for |H| = infinity

- Using $VC(H)$ as a measure of expressiveness, we can get an Occam algorithm for infinite hypothesis spaces.

- Given a sample D of m examples, find some $h \in H$ that is consistent with all m examples
- If $m > \frac{1}{\varepsilon}\{8VC(H)\log\frac{13}{\varepsilon} + 4\log\left(\frac{2}{\delta}\right)\}$

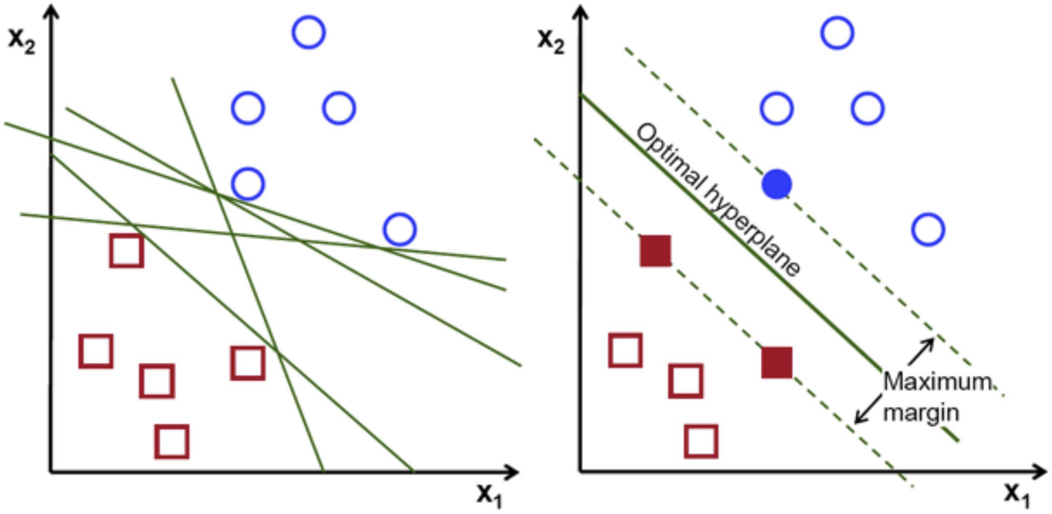
- Then with probability at least $(1-\delta)$, h has error less than ε. (that is, if m is polynomial we have a PAC learning algorithm; to be efficient, we need to produce the hypothesis h efficiently.

Penn Engineering

Part 2: SVM

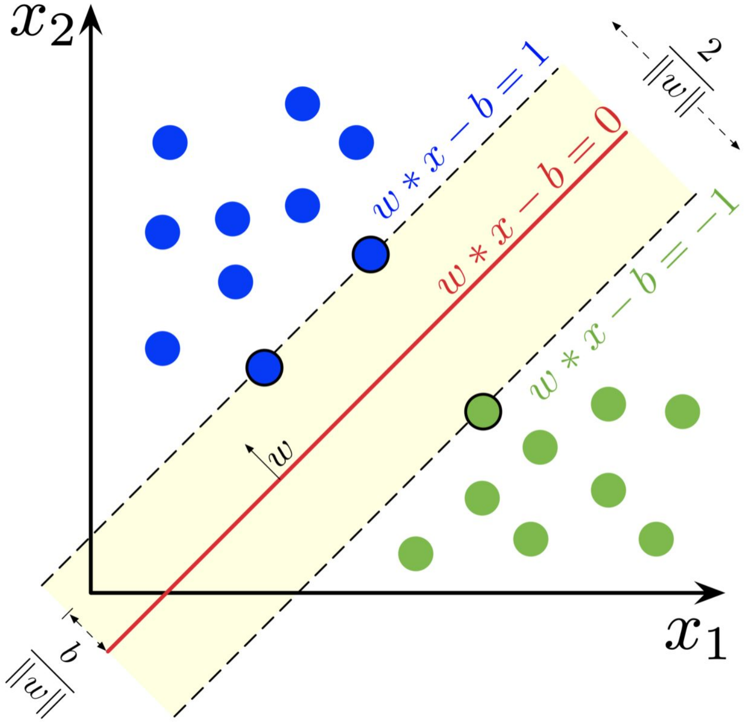
What is Support Vector Machine?

- Finding the separator that maximize the margin of 2 sets of data



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

Margin

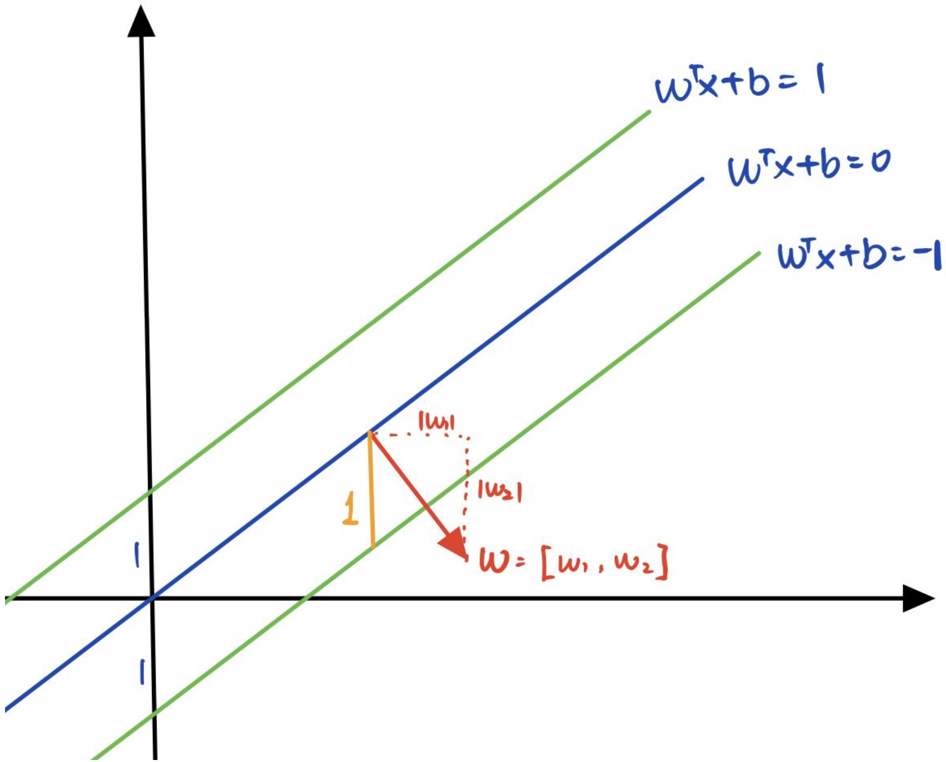


$$\min_{w,b} \; \frac{1}{2}\boldsymbol{w}^T\boldsymbol{w}$$

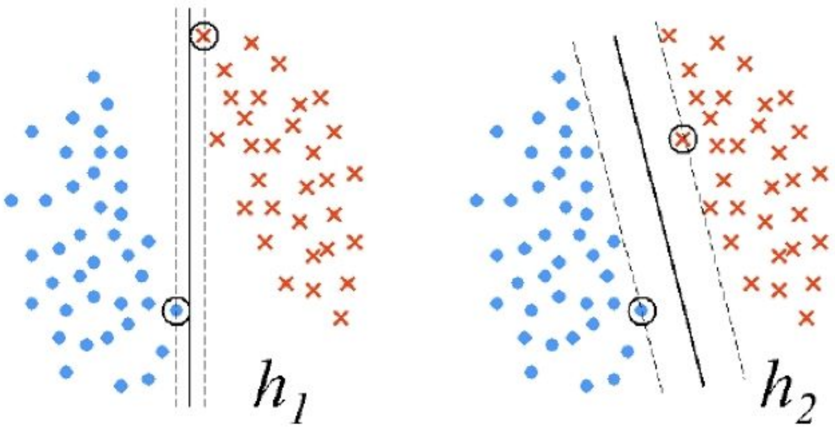
$$\text{s.t} \quad y_i(\mathbf{w}^{\mathrm{T}}\boldsymbol{x}_i + b) \geq 1, \forall(\boldsymbol{x}_i, y_i) \in S$$

- Why Margin = 1/||w||?
- Where does it come from?

Margin-Geometry Perspective



Hard SVM



The margin of a linear separator $\boldsymbol{w}^T \boldsymbol{x} + b = 0$ is $\frac{1}{||\boldsymbol{w}||}$

$$\max \frac{1}{||\boldsymbol{w}||} = \min ||\boldsymbol{w}||$$
$$= \min \tfrac{1}{2} \boldsymbol{w}^T \boldsymbol{w}$$

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

$$\text{s.t} \quad y_i(\mathbf{w}^{\mathrm{T}} \boldsymbol{x}_i + b) \geq 1, \forall (\boldsymbol{x}_i, y_i) \in S$$

Soft SVM

- The problem we solved is:

$$Min\ ½\ ||\boldsymbol{w}||^2 + c\sum \xi_i$$

- Where $\xi_i > 0$ is called **a slack variable**, and is defined by:
  - $\xi_i = \max(0, 1 - y_i \boldsymbol{w}^T \boldsymbol{x}_i)$
  - Equivalently, we can say that: $y_i \boldsymbol{w}^T \boldsymbol{x}_i \geq 1 - \xi_i;\ \xi_i \geq 0$
- And this can be written as:

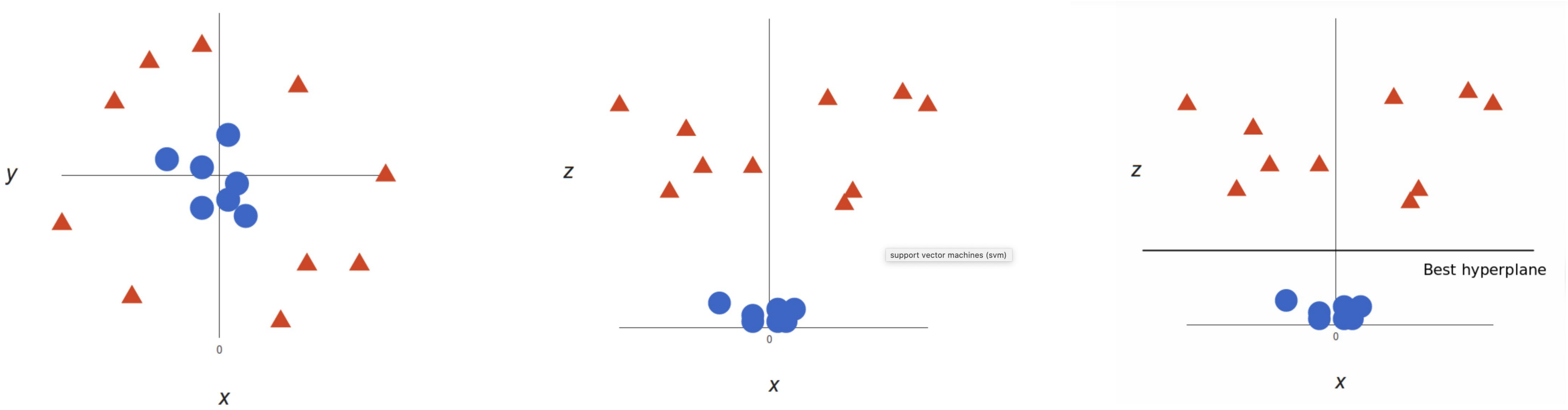
$$\underbrace{Min\ ½\ ||\boldsymbol{w}||^2}_{\text{Regularization term}} + \underbrace{c\sum \xi_i}_{\text{Empirical loss}}$$

Can be replaced by other **regularization functions**

Can be replaced by other **loss functions**

- General Form of a learning algorithm:
  - Minimize empirical loss, and Regularize (to avoid over fitting)
  - Theoretically motivated improvement over the original algorithm we've seen at the beginning of the semester.

Nonlinear SVM & Kernels



Part 3: Adaboost

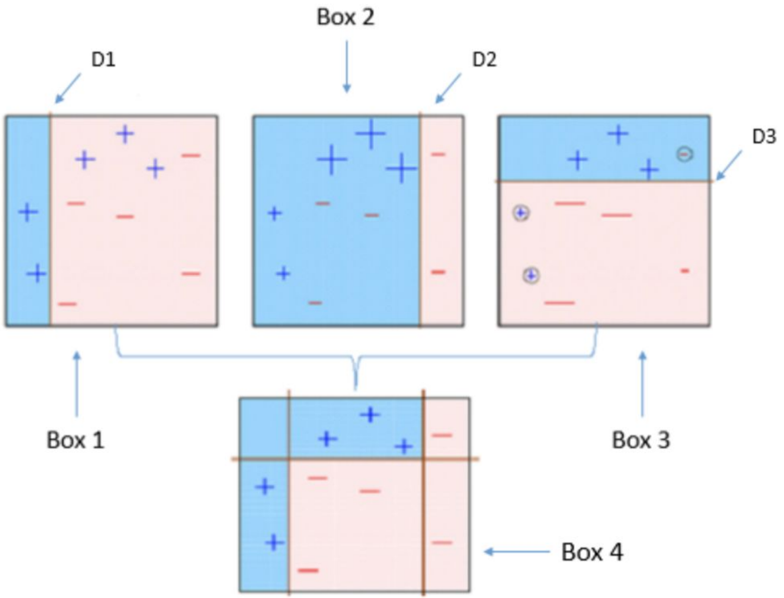
More Intuition

What is Adaboost?

- A predictors which trains sequentially, each trying to correct its predecessor.

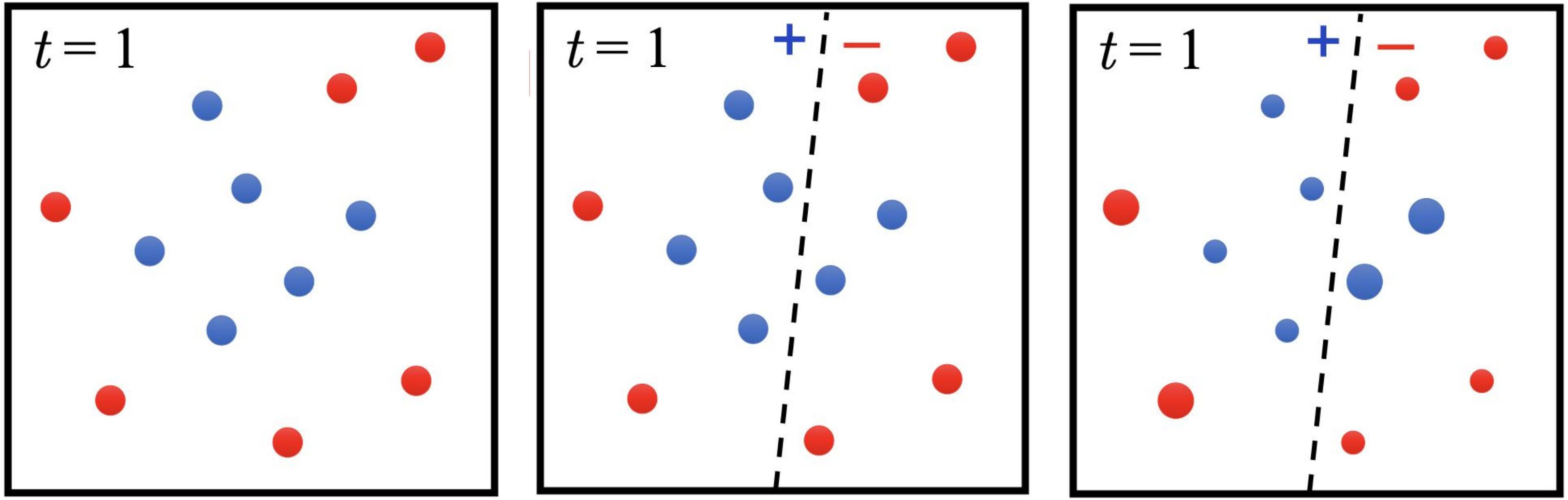
- **Method:** set weights to both classifiers and data points in a way that forces classifiers to concentrate on observations that are difficult to correctly classify.

- Helps combine multiple **"weak classifiers"** into **a single "strong classifier"**



Adaboost with DT Stump

Adaboost with Linear Separator



Adaboost with Linear Separator

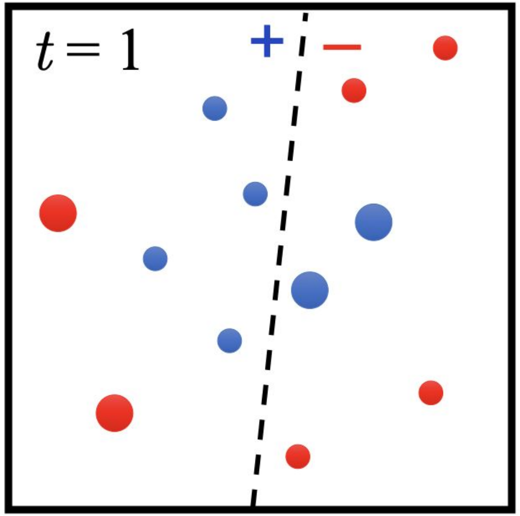
- Constructing D_t on $\{1, \ldots m\}$:
  - $D_1(i) = 1/m$
  - Given D_t and h_t :
    - $D_{t+1} = D_t(i)/z_t \times e^{-\alpha_t}$ if $y_i = h_t(\boldsymbol{x}_i)$

      $D_t(i)/z_t \times e^{+\alpha_t}$ if $y_i \neq h_t(\boldsymbol{x}_i)$

      $$\boxed{= \frac{D_t(i)}{z_t} \times \exp(-\alpha_t \, y_i \, h_t(\boldsymbol{x}_i))}$$

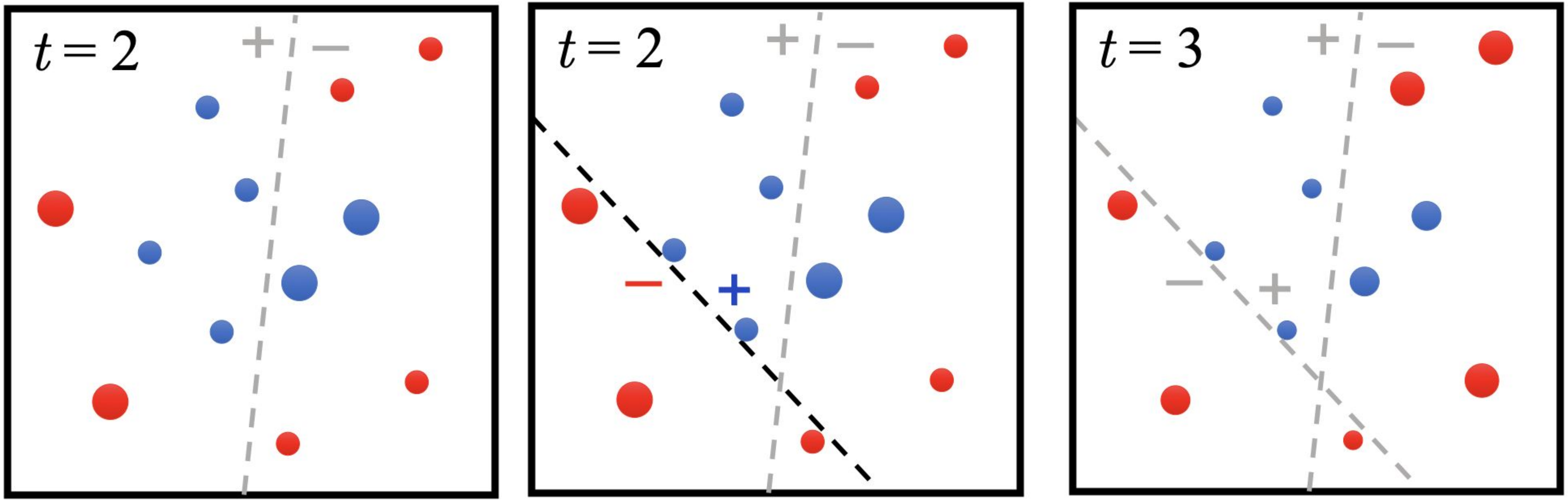
    where z_t = normalization constant

    and $\alpha_t = ½ \ln\{(1 - \varepsilon_t)/\varepsilon_t\}$



- Final hypothesis: $H_{final}(\boldsymbol{x}) = sign(\sum_t \alpha_t h_t(\boldsymbol{x}))$

Adaboost with Linear Separator



Adaboost with Linear Separator