Linear Machines and SVM – Part 3: SVM for Linearly Separable Case

Ira A. Fulton Schools of Engineering
Arizona State University

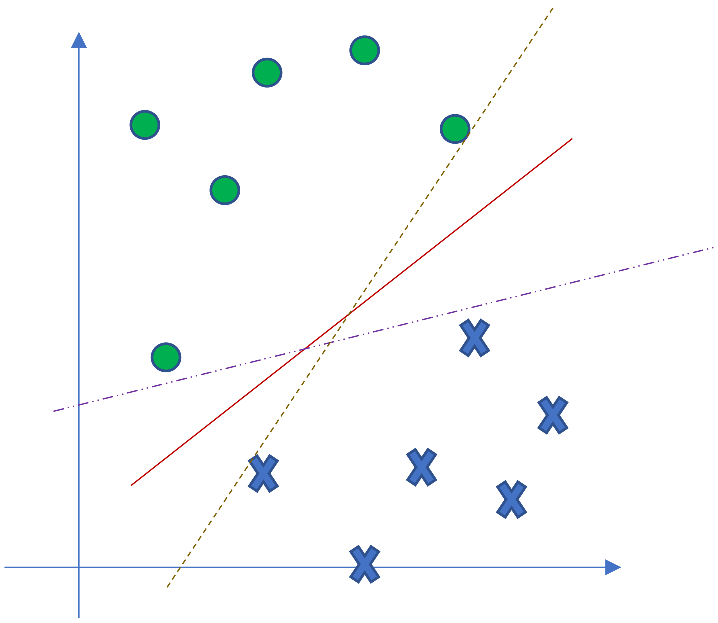
Objective

Objective

Construct SVM for
Linearly Separable Data

Key Idea of Support Vector Machines

- For a given set, a classifier that gives rise to a larger margin will be better.

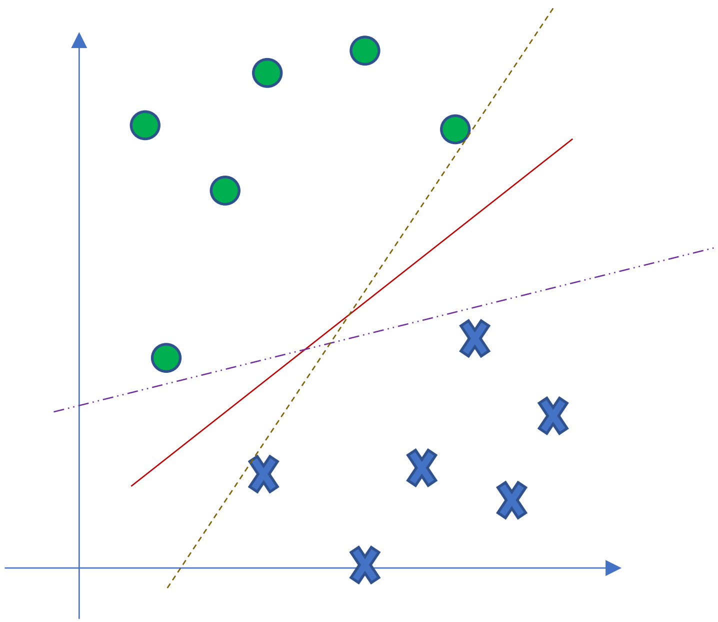


- SVM: To find the decision boundary such that the margin is maximized.

Formulating the Problem

- Given labeled training data:

$<\mathbf{x}^{(i)}, y^{(i)}>$, $y^{(i)} \in \{-1,1\}$, $\mathbf{x}^{(i)} \in \boldsymbol{R}^{d}$, $i=1,\ldots,n$,



- Assuming the points are linearly separable, let's write a separating hyperplane as:

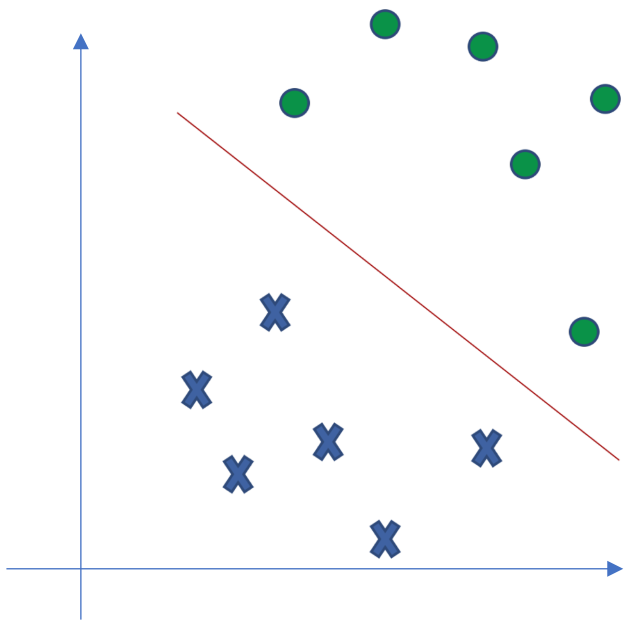
$$H: \quad \mathbf{w}^{t}\mathbf{x} + b = 0$$

Formulating the Problem (cont'd)

- Let d_+ (d_-) be the shortest distance from the separating hyperplane to the *closest* positive (negative) examples.

- These defines planes H_1 and H_2.

- We can let $d_+ = d_- = d$
  - ➔ Find a solution maximizing 2d.

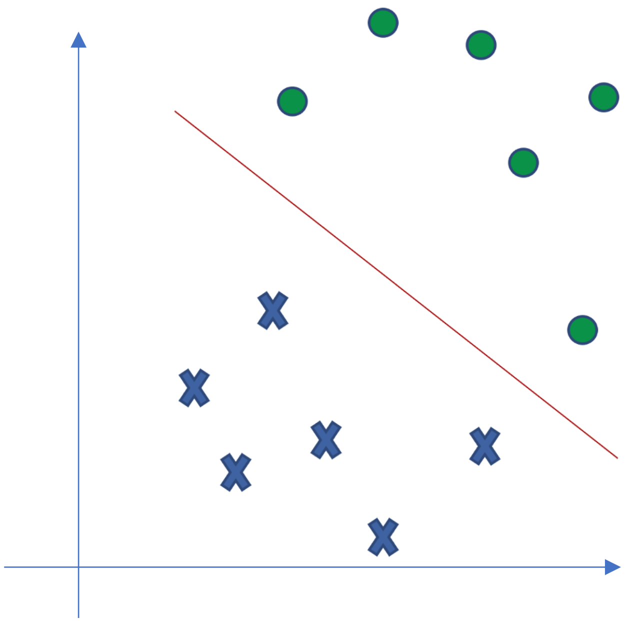


Formulating the Margin

- Given separating plane H: $\mathbf{w}^t\mathbf{x} + b = 0$ and distance d,

  what are the equations for H_1 and H_2?

- Consider the plane H*
  given by $\mathbf{w}^t\mathbf{x} + b = ||\mathbf{w}||d$
  - Check its orientation
  - Check its distance to H



Formulating the Margin (cont'd)

- H_1 is given by $\mathbf{w}^t\mathbf{x} + b = ||\mathbf{w}||d$

- Similarly, H_2 is given by $\mathbf{w}^t\mathbf{x} + b = -||\mathbf{w}||d$

- Note: for any plane equation, $\mathbf{w}^t\mathbf{x} + b = 0$, $\{\mathbf{w}, b\}$ is defined only up to an unknow scale:
  - $\{s\mathbf{w}, sb\}$ is also a valid solution to the equation, for any constant s.

Formulating the Margin (cont'd)

- We can have the canonical formulation for all the planes as

H: $\mathbf{w}^t\mathbf{x} + b = 0$

H_1: $\mathbf{w}^t\mathbf{x} + b = 1$

H_2: $\mathbf{w}^t\mathbf{x} + b = -1$

- The region between H_1 and H_2 is also called the margin, and its width is $\frac{2}{||\mathbf{w}||}$

Formulating SVM

$$\{\mathbf{w}^*, b^*\} = \underset{\mathbf{w},b}{\operatorname{argmin}} \|\mathbf{w}\| \quad or \quad \{\mathbf{w}^*, b^*\} = \underset{\mathbf{w},b}{\operatorname{argmin}} \tfrac{1}{2}\|\mathbf{w}\|^2$$

Subject to

$$\mathbf{w}^t\mathbf{x}^{(i)} + b \geq 1 \quad \text{for } y^{(i)} = +1$$

$$\mathbf{w}^t\mathbf{x}^{(i)} + b \leq -1 \quad \text{for } y^{(i)} = -1$$

The constraints can be combined into:

$$y^{(i)}(\mathbf{w}^t\mathbf{x}^{(i)} + b) - 1 \geq 0 \quad \forall i$$

➔ A nonlinear (quadratic) optimization problem with linear inequality constraints.

How to solve SVM? (Outline)

- Reformulate the problem using Lagrange multipliers α
  - Lagrangian Primal Problem
  - Lagrangian Dual Problem

- The Karush-Kuhn-Tucker Conditions
  - *Necessary and sufficient* for $\mathbf{w}$, b, α.
  - Solving the SVM problem ➔ finding a solution to the KKT conditions.

SVM: Lagrangian Primal Formulation

Define

$$L_P(\mathbf{w}, b, \alpha) = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_i \alpha_i[y^{(i)}(\mathbf{w}^t\mathbf{x}^{(i)} + b) - 1]$$

then the SVM solution
should satisfy

$$\frac{\partial L_P}{\partial \mathbf{w}} = 0, \qquad \frac{\partial L_P}{\partial \mathrm{b}} = 0,$$

$$\alpha_i \geq 0,$$

$$\alpha_i[y^{(i)}(\mathbf{w}^t\mathbf{x}^{(i)} + b) - 1]=0$$

$\rightarrow$

The final **w** is given by

$$\mathbf{w} = \sum_i \alpha_i y^{(i)} \mathbf{x}^{(i)}$$

and b is given by

$$y^{(k)} - \mathbf{w}^t\mathbf{x}^{(k)}$$

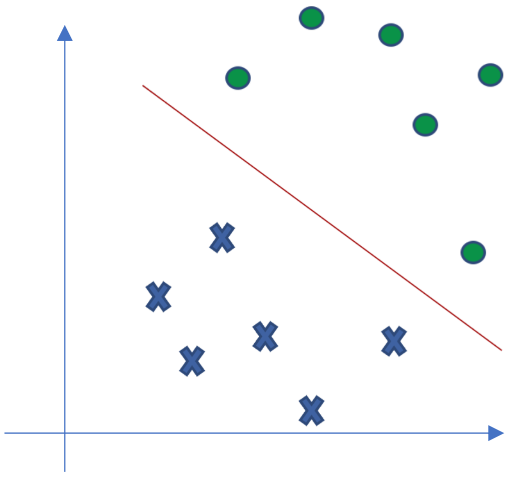
for any k such that $\alpha_k > 0$

SVM: Lagrangian Dual Formulation

The objective function is

$$L_D(\mathbf{w}, b, \alpha) = \sum_i \alpha_i - \frac{1}{2} \sum_{i,j} \alpha_i \alpha_j y^{(i)} y^{(j)} \underline{\mathbf{x}^{(i)} \cdot \mathbf{x}^{(j)}}$$

The solution is the same as before. But there is an important observation.



Points for which $\alpha_i > 0$ are called **support vectors**