Support Vector Machines (I)

COMP4211

THE DEPARTMENT OF
COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

Classification Problem

Given: Training set $\{(\mathbf{x}_1, y_1), (\mathbf{x}_2, y_2), \ldots, (\mathbf{x}_N, y_N)\}$

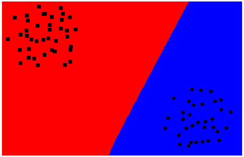
$(\mathbf{x}_i, y_i)$: training pattern

$\mathbf{x}_i \in \mathbb{R}^m$: input

$y_i \in \{\pm 1\}$: output (label) (in general, can have > 2 classes)

Assume that the problem is linearly separable

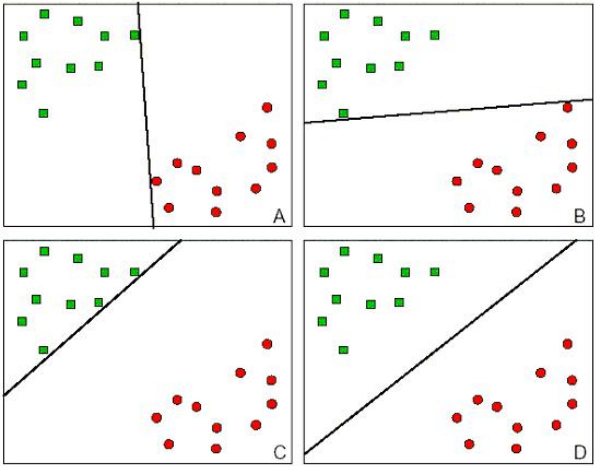
- there exists a linear surface to separate the two classes



- 2-D: line; 3-D: plane; ...; n-D: hyperplane $(\mathbf{w}'\mathbf{x} + b = 0)$

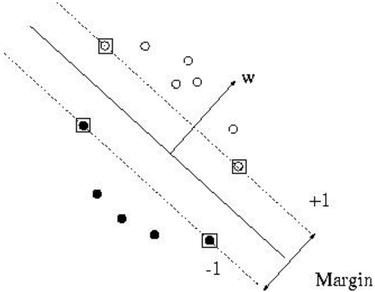
Find $\mathbf{w}'\mathbf{x} + b = 0$ that perfectly separates the two classes

Multiple Solutions



“Optimal” Hyperplane

Idea: maximize the margin



$$\mathbf{w}'\mathbf{x} + b = \begin{cases} 1 & \text{for the closest points on one side} \\ -1 & \text{for the closest points on the other} \end{cases}$$

$$\mathbf{w}'\mathbf{x}_1 + b = 1 \quad \mathbf{w}'\mathbf{x}_2 + b = -1$$

Formulation as Optimization Problem

$$\text{margin} = \frac{\mathbf{w}}{\|\mathbf{w}\|}'(\mathbf{x}_1 - \mathbf{x}_2) = \frac{2}{\|\mathbf{w}\|}$$

the hyperplane should separate the two classes

$$\mathbf{w}'\mathbf{x} + b \begin{cases} \geq 1 & \text{if } y_i = 1 \\ \leq -1 & \text{if } y_i = -1 \end{cases}$$

- or, equivalently, $y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1$

$$\min \ \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to} \ \ y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1, \quad \forall i$$

- constrained optimization problem

(*) Method of Lagrange Multipliers

$$\begin{array}{ll} \text{minimize} & f_0(\mathbf{x}) \\ \text{subject to} & h_i(\mathbf{x}) = 0, \; i = 1, \ldots, p \end{array}$$

define Lagrangian $\mathcal{L} : \mathbb{R}^{n+p} \to \mathbb{R}$

$$\mathcal{L}(\mathbf{x}, \boldsymbol{\nu}) = f_0(\mathbf{x}) + \nu_1 h_1(\mathbf{x}) + \ldots + \nu_p h_p(\mathbf{x})$$

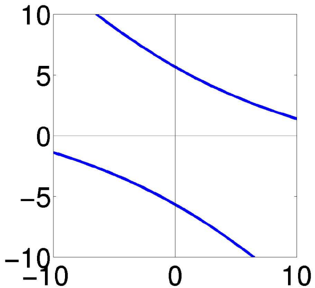
- ν_i: Lagrange multipliers or dual variables
- objective is augmented with weighted sum of constraint functions

at optimality

$$\begin{aligned} h_i(\mathbf{x}^*) &= 0 \\ \nabla f_0(\mathbf{x}^*) + \sum_i \nu_i^* \nabla h_i(\mathbf{x}^*) &= 0 \end{aligned}$$

Example

Find the shortest distance from the origin to the hyperbola $x^2 + 8xy + 7y^2 = 225$



Find the minimum value of $x^2 + y^2$ subject to the constraint $x^2 + 8xy + 7y^2 - 225 = 0$

(*) Solution

Find the minimum value of $x^2 + y^2$ subject to the constraint $x^2 + 8xy + 7y^2 - 225 = 0$

$$\mathcal{L}(x, y) = x^2 + y^2 - \lambda(x^2 + 8xy + 7y^2 - 225)$$

$$\frac{\partial L}{\partial x} = 2x - \lambda(2x + 8y) = 0$$

$$\frac{\partial L}{\partial y} = 2y - \lambda(8x + 14y) = 0$$

$(x, y) \neq (0, 0)$

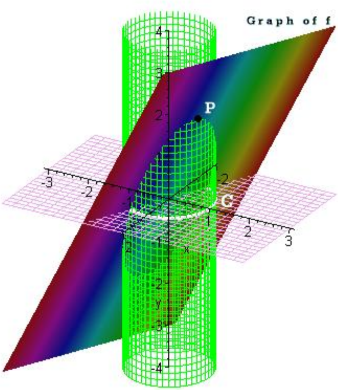
$$\begin{aligned} \begin{vmatrix} 1-\lambda & -4\lambda \\ -4\lambda & 1-7\lambda \end{vmatrix} &= 0 \\ 9\lambda^2 + 8\lambda - 1 &= 0 \end{aligned}$$

$$\lambda = 1/9 \quad \text{or} \quad \lambda = -1$$

1. $\lambda = -1$
   - plug into the equations $\frac{\partial L}{\partial x} = 0$ and $\frac{\partial L}{\partial y} = 0$
   - $-5y^2 = 225$
   - no solution
2. $\lambda = 1/9$
   - $2x = y$
   - substitute into $x^2 + 8xy + 7y^2 = 225$
   - $x^2 = 5, y^2 = 20, x^2 + y^2 = 25$
   - distance=5

(*) Another Example

Find the maximum and minimum values of $f(x, y) = x + 2y$ subject to the constraints $x^2 + y^2 = 1$



(*) Solution

$$\mathcal{L}(x, y) = x + 2y - \lambda(x^2 + y^2 - 1)$$

$$\frac{\partial L}{\partial x} = 1 - \lambda(2x) = 0, \quad \frac{\partial L}{\partial y} = 2 - \lambda(2y) = 0, \quad x^2 + y^2 = 1$$

- $\lambda, x, y \neq 0$; $y = 2x$; $x^2 + 4x^2 = 1$
- $x = \pm\frac{1}{\sqrt{5}}; y = \pm\frac{2}{\sqrt{5}}$

$$f\left(\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right) = \sqrt{5} \qquad f\left(\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}}\right) = -\frac{3}{\sqrt{5}}$$

$$f\left(-\frac{1}{\sqrt{5}}, \frac{2}{\sqrt{5}}\right) = \frac{3}{\sqrt{5}} \qquad f\left(\frac{1}{\sqrt{5}}, -\frac{2}{\sqrt{5}}\right) = -\sqrt{5}$$

Back to ML

$$\text{(primal)} \quad \min \quad \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{s.t.} \quad y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1, \quad \forall i$$

Method of Lagrange multipliers

- associate one Lagrange multiplier α_i with each constraint

$$\text{(dual)} \quad \max \quad \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}_i' \mathbf{x}_j$$
$$\text{s.t.} \quad \alpha_i \geq 0$$
$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

(*) Proof

$$\text{minimize } \frac{1}{2}\|\mathbf{w}\|^2 \quad \text{subject to } \; y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1, \quad \forall i$$

Lagrangian: $\mathcal{L} = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{N} \alpha_i (y_i(\mathbf{w}'\mathbf{x}_i + b) - 1)$

$$\nabla_{\mathbf{w},b}\mathcal{L} = 0 \Rightarrow \begin{cases} \frac{\partial L}{\partial \mathbf{w}} = 0 & \Rightarrow \quad \mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i \\ \frac{\partial \mathcal{L}}{\partial b} = 0 & \Rightarrow \quad \sum_{i=1}^{N} \alpha_i y_i = 0 \end{cases}$$

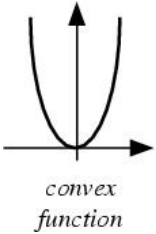
Substitute back in the primal to get the dual

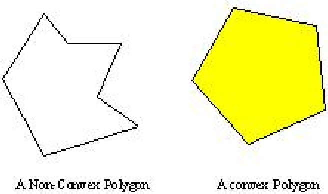
$$\text{maximize} \quad \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}_i' \mathbf{x}_j$$

$$\text{subject to} \quad \alpha_i \geq 0, \quad \sum_{i=1}^{N} \alpha_i y_i = 0$$

(*) Convex Programming

minimize a convex function on a convex set





(*) Examples

Linear programming

- linear objective function, linear constraints

  minimize $'\mathbf{x}$ subject to $\mathbf{a}_i'\mathbf{x} - b_i \leq 0, \quad i = 1, \ldots, m$

Quadratic programming (QP)

- quadratic objective function, linear constraints

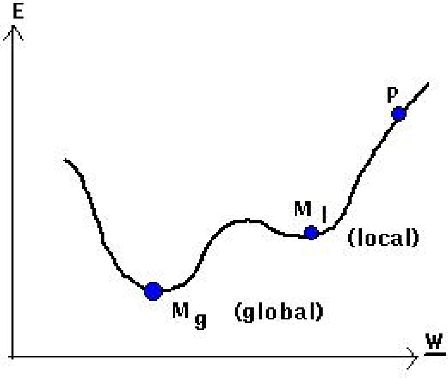
$$
\begin{array}{ll}
\min & \frac{1}{2}\mathbf{x}'\mathbf{G}\mathbf{x} + \mathbf{g}'\mathbf{x} \\
\text{subject to} & \left\{ \begin{array}{l} \mathbf{a}_i'\mathbf{x} = b_i \\ \mathbf{a}_i'\mathbf{x} \geq b_i \end{array} \right.
\end{array}
$$

- $\mathbf{G}$: (positive semi-definite) matrix; $\mathbf{g}$: vector

(*) Global Optimality

Every local solution is a global solution

- does not have the problem of local optimum



Back to ML

$$\text{(primal)} \quad \min \quad \frac{1}{2}\|\mathbf{w}\|^2$$
$$\text{s.t.} \quad y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1, \quad \forall i$$

$$\text{(dual)} \quad \max \quad \sum_{i=1}^{N} \alpha_i - \frac{1}{2} \sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}_i' \mathbf{x}_j$$
$$\text{s.t.} \quad \alpha_i \geq 0$$
$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

- quadratic programming problem
- can be solved numerically by any general purpose optimization packages, e.g. MATLAB optimization toolbox

Support Vectors

- Patterns for which $y_i(\mathbf{w}'\mathbf{x}_i + b) > 1$
  - can be shown that $\alpha_i = 0$
  - $\mathbf{x}_i$ irrelevant
- Patterns that have $\alpha_i > 0$
  - can be shown that $y_i(\mathbf{w}'\mathbf{x}_i + b) = 1$
  - lie either on H_1 or H_2
- Solution is determined by the examples on the margin (support vectors)
- If all other training points are removed or moved around, and training was repeated, the same hyperplane would be found

How to Find b?

- find any support vector $\mathbf{x}^*(1)$ that belongs to the first class

$$\mathbf{w}'\mathbf{x}^*(1) + b = 1$$

- find any support vector $\mathbf{x}^*(-1)$ that belongs to the second class

$$\mathbf{w}'\mathbf{x}^*(-1) + b = \ -1$$

$$b = -\frac{1}{2}(\mathbf{w}'\mathbf{x}^*(1) + \mathbf{w}'\mathbf{x}^*(-1))$$

Typically, better perform averaging over all SV's

How to Perform Testing?

- use $\mathbf{w}'\mathbf{x} + b$
- recall that $\mathbf{w} = \sum_{i=1}^{N} \alpha_i y_i \mathbf{x}_i$

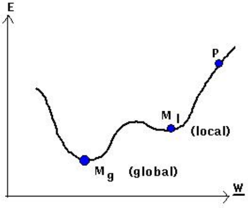
$$\begin{aligned} \text{sign}(\mathbf{w}'\mathbf{x} + b) &= \text{sign}\left(\sum_{i=1}^{N} \alpha_i y_i {\mathbf{x}_i}'\mathbf{x} + b\right) \\ &= \text{sign}\left(\sum_{i=1}^{N_S} \alpha_i y_i \mathbf{x}_i'\mathbf{x} + b\right) \end{aligned}$$

- N_S: number of support vectors

Note

Recall that every local solution to a convex programming problem is a globally optimal solution

- contrast to neural networks, where many local minima usually exist



In both training and testing, training data only appear in the form of dot products between vectors

- will become important later on

What if Training Data not Linearly Separable?

separate the training set with a minimal number of errors

introduce positive slack variables ξ_i's ($\xi_i \geq 0$)

$$\begin{cases} \mathbf{w}'\mathbf{x}_i + b \geq 1 - \xi_i & y_i = 1 \\ \mathbf{w}'\mathbf{x}_i + b \leq -1 + \xi_i & y_i = -1 \end{cases}$$

penalize $\sum_i \xi_i$ in the objective function

$$\min \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i$$

$$\text{s.t.} \quad y_i(\mathbf{w}'\mathbf{x}_i + b) \geq 1 - \xi_i, \xi_i \geq 0$$

- soft margin hyperplane

$$(\text{dual}) \quad \max \quad \sum_{i=1}^{N} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{N} \alpha_i \alpha_j y_i y_j \mathbf{x}_i'\mathbf{x}_j$$

$$\text{s.t.} \quad C \geq \alpha_i \geq 0, \quad i = 1, \ldots, N$$

$$\sum_{i=1}^{N} \alpha_i y_i = 0$$

- still a QP problem $\rightarrow$ every solution is a global solution