Convex Optimization
CMU-10725

2. Linear Programs

Barnabás Póczos & Ryan Tibshirani

MACHINE LEARNING DEPARTMENT
Carnegie Mellon.
School of Computer Science

Administrivia

- **Please ask questions!**
- Lecture = 40 minutes part 1 - 5 minutes break – 35 minutes part 2
- Slides: http://www.stat.cmu.edu/~ryantibs/convexopt/
- Anonym feedback survey will be on black board next week. **Please use it!** Constructive feedback and suggestions are always welcome!
- Subscribe for scribing!
- My office hour is after the class.

Basic Definitions

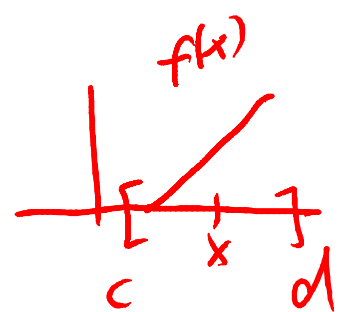
- More and more complicated optimization problems
- Definition of LP

Simplest Optimization Problems

Goal: $\min f(x)$ or $\max f(x)$

- Constant function $f(x) = c$
- 1-dim linear function

  $f(x) = ax + b$
- 1-dim linear function with bound constraints



Linear Programs

- n-dim linear function with m linear constraints

Inequality form:

Cost function: $c_1 x_1 + c_2 x_2 + \dots + c_n x_n$

Constraints:

s.t. $a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n \le b_1$

$\vdots$

$a_{m1} x_1 + a_{m2} x_2 + \dots + a_{mn} x_n \le b_m$

Bounds: $l_1 \le x_1 \le u_1$ $\quad l_i \quad -\infty$

$\vdots \qquad\qquad\qquad\qquad \infty$

$l_n \le x_n \le u_n$

Linear Programs

Inequality form using matrix notation:

Cost function: $c^T x \quad x \in \mathbb{R}^n$

Constraints: $Ax \le b \quad b \in \mathbb{R}^m \quad A^T \in \mathbb{R}^{n \times m}$

Bounds: $l \le x \le u$

Example:

$$\min \; -2x_1 - x_2$$

$$\text{s.t.} \quad x_1 + x_2 \le 5$$

$$2x_1 + 3x_2 \le 4$$

$$x_1 \ge 0 \quad x_2 \ge 0$$

$$c = [-2, -1]^T$$

$$b = [5, 4]^T$$

$$A = \begin{bmatrix} 1 & 1 \\ 2 & 3 \end{bmatrix}$$

Goal of this (...and next) lecture(s)

- **To be able to solve Linear Programs**

  Simplex Algorithm (Phase I and Phase II)
  (Later we will see other algorithms too)

- **Understand why LP is useful**
  - Motivation
  - Applications in Machine Learning

- **Understand the difficulties**
  - Convergence? Polynomial or Exponential many operations?
  - Will algorithms find the exact solutions, or only approximate ones?

Table of Contents

- **Motivating Examples & Applications:**
  - Pattern classification
- **Linear programs**:
  - standard form
  - canonical form
- **Solutions:**
  - Basic, Feasible, Optimal, Degenerate
- **Simplex algorithm:**
  - Phase I
  - Phase II

Linear Programs

- Motivation
- History
- Sketching LP

History

Dantzig 1947 (Simplex method)

(one of the top 10 algorithms of the twentieth century**)**

Motivated by World War II:

- Job scheduling (Assign 70 men to 70 jobs)
- Blending problem
  (produce a blend (30% Lead, 30% Zinc, 40% Tin) out of 9 different alloys (different mixture, different costs) such that the cost is minimal)
- Network flow optimization (Max flow min cut)

The product mix problem

A furniture company manufactures four models of desks

Number of man hours and profit:

	Desk 1	Desk 2	Desk 3	Desk 4	Available hrs
Carpentry shop hrs	4	9	7	10	6000
Finishing shop hrs	1	1	3	40	4000
Profit	\$12	\$20	\$18	\$40	

$$x_1 \geq 0 \quad x_2 \geq 0, \quad x_3 \geq 0, \quad x_4 \geq 0, \quad x_5 \geq 0, \quad x_6 \geq 0$$

$$\text{MAX PROFIT} = 12x_1 + 20x_2 + 18x_3 + 40x_4$$

$$\text{s.t.} \quad 4x_1 + 9x_2 + 7x_3 + 10x_4 + x_5 \leq 6000$$

$$x_1 + x_2 + 3x_3 + 40x_4 + x_6 \leq 4000$$

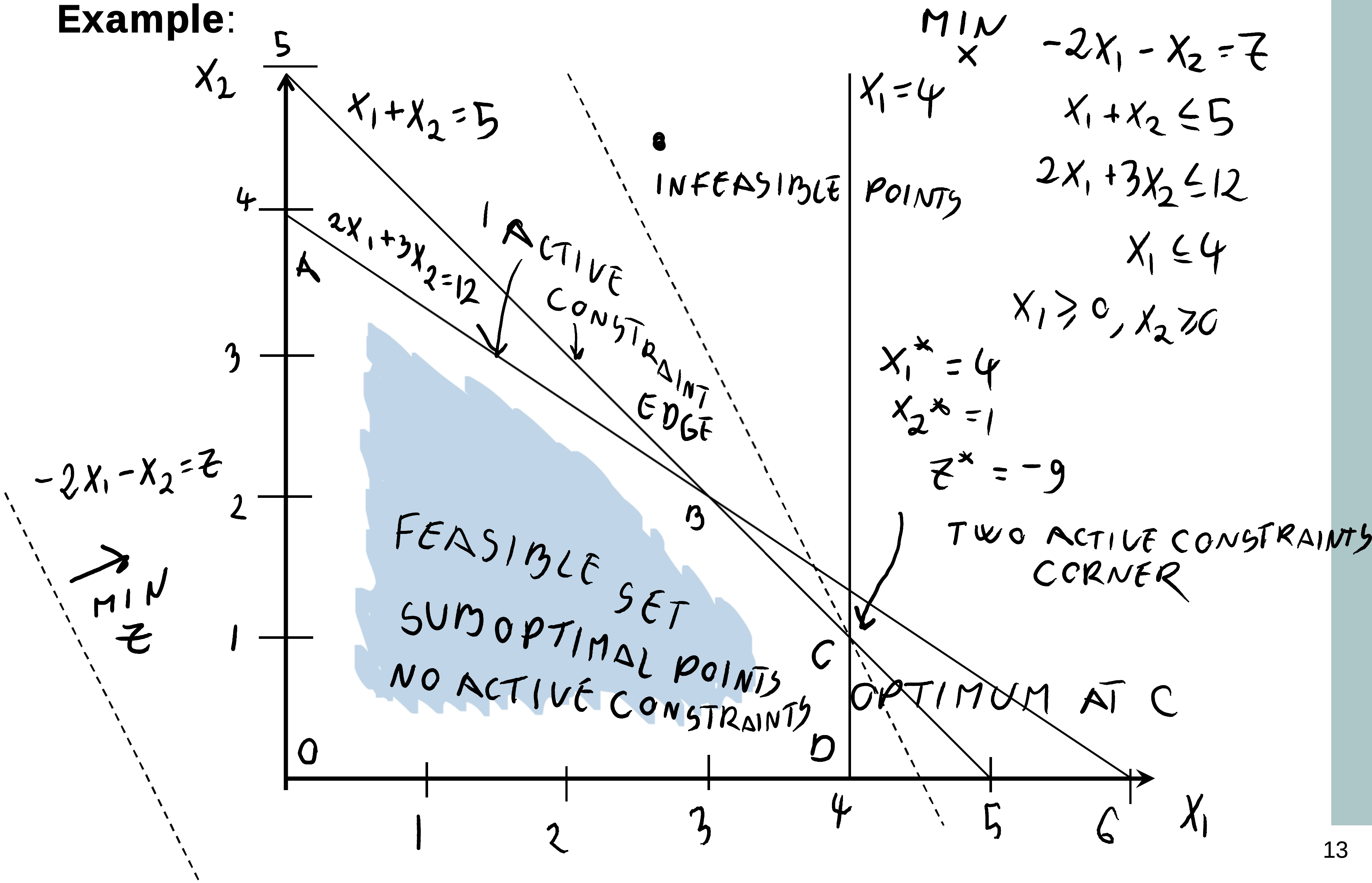
Why is it called Linear Programing???

Motivation: Why Linear Programing?

- The simplest, nontrivial optimization problem
- Many complex system (objective and constraints) can be well approximated with linear equations
- Important applications
- There are efficient toolboxes that can solve LPs

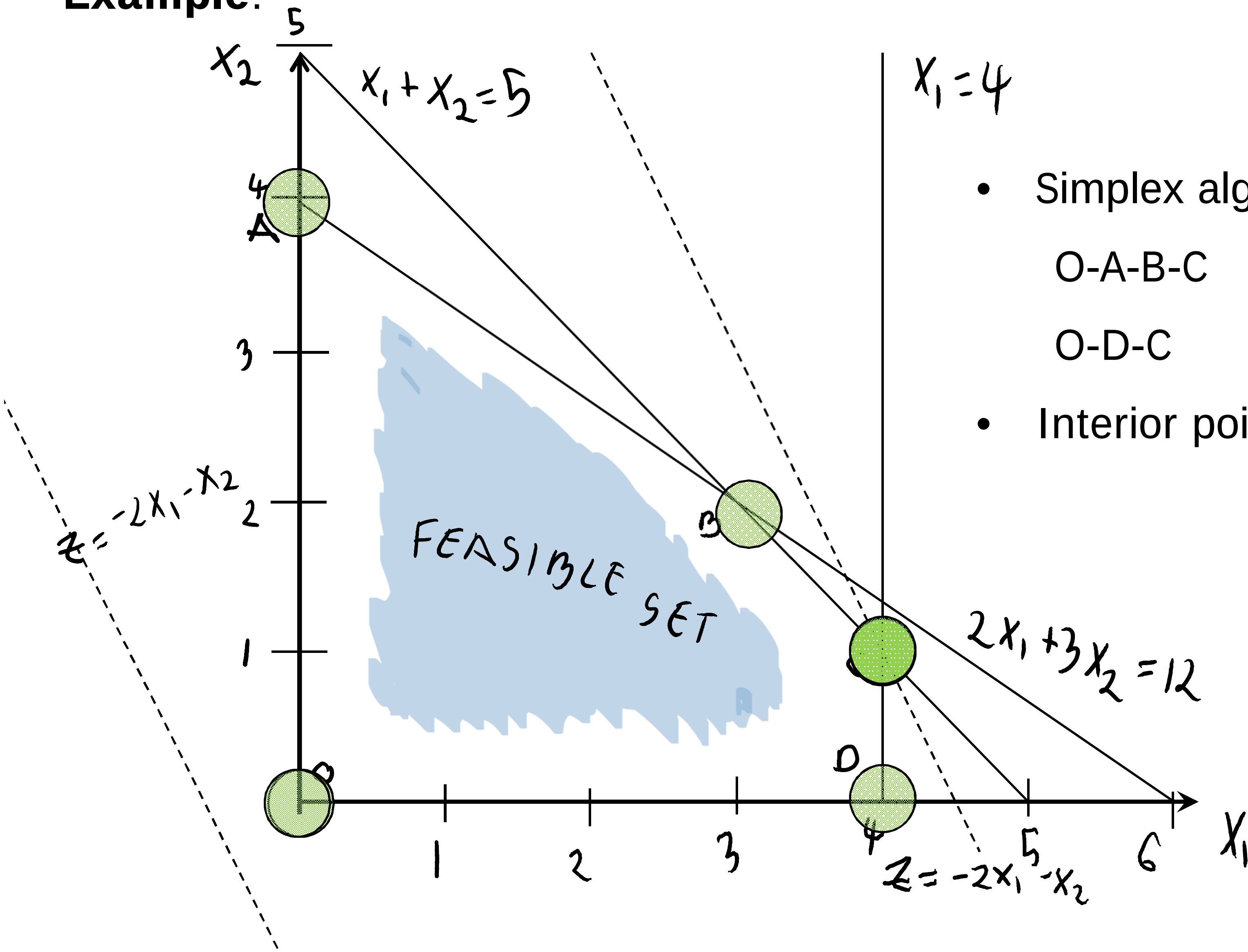
Sketching Linear Programs

Example:



Simplex Algorithm

Example:

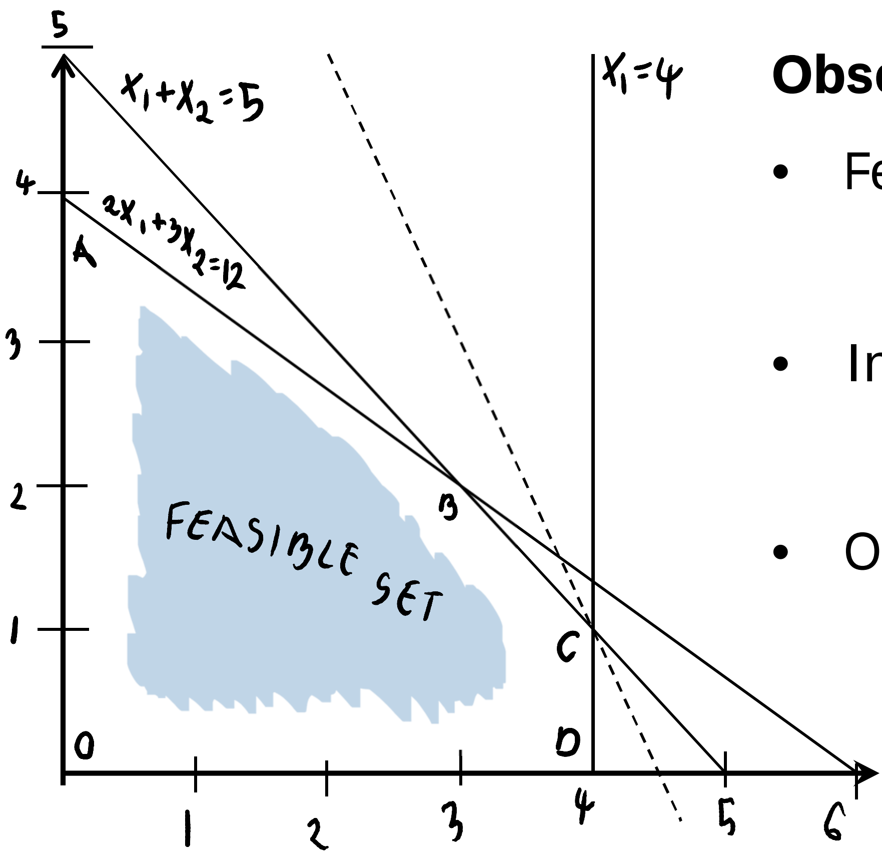


- Simplex algorithm:

  O-A-B-C

  O-D-C
- Interior point methods

Linear Program



$$\text{MIN [OR MAX] } c^T x \qquad c \in \mathbb{R}^n$$
$$\text{S.T. } Ax \le b \qquad A \in \mathbb{R}^{m \times n}$$
$$l \le x \le u \qquad l, u, x \in \mathbb{R}^n$$

Observations, Difficulties:

- Feasible set might not exist, no solution

  (Inconsistency in the constraints)

- Infinite many global optimum

  (Optimum is on an edge)

- Optimum can be – 1 , 1

  (Unbounded optimum)

Linear Program

High dimensional case is similar:

faces, facets instead of edges

cost function = hyperplane

MIN [OR MAX] $c^T x$ $\quad c \in \mathbb{R}^n$

S.T. $Ax \le b$ $\quad A \in \mathbb{R}^{m \times n}$

$\ell \le x \le u$ $\quad \ell, u, x \in \mathbb{R}^n$

Applications

Pattern Classification via Linear Programming

Application

Pattern Classification via Linear Programming

More info can be found on: cgm.cs.mcgill.ca/~beezer/cs644/main.html

Goal: show how LP can be used for linear classification.

Why LP?

There are many efficient LP solver software packages

Pattern Classification via LP

Formal goal:

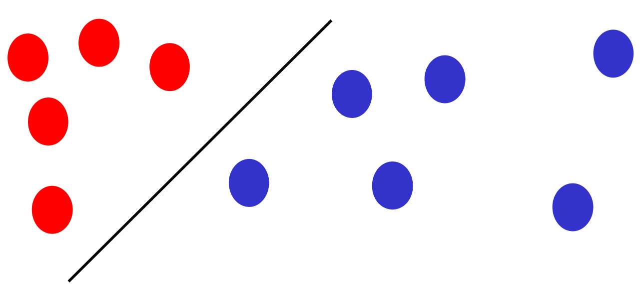
GIVEN $H = \{H^1, H^2, \ldots, H^h\} \subseteq \mathbb{R}^n$

$M = \{M^1, M^2, \ldots M^m\} \subseteq \mathbb{R}^n$

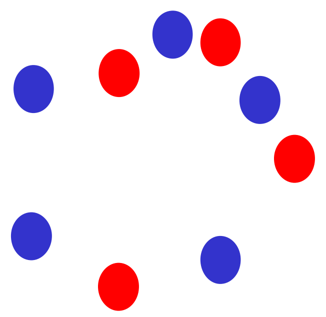
Problem 1: Determine whether H and M are linearly separable

Problem 2: If H and M are linearly separable,
then find a separating hyper plane

Linearly separable sets:

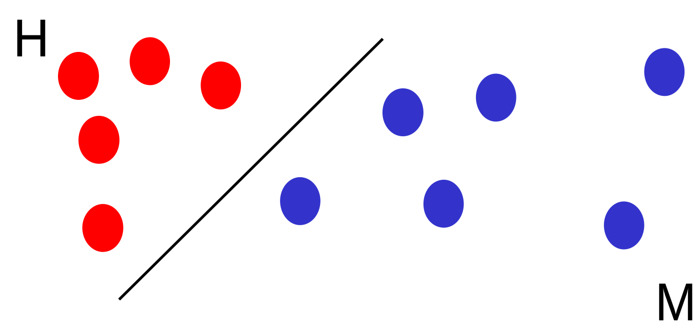


Linearly not separable sets:

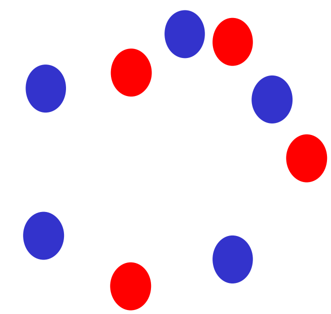


Pattern Classification via LP

Linearly separable sets:



Linearly not separable sets:



Observation:

H and M are linearly separable $\iff \exists\, a \in \mathbb{R}^n,\ b \in \mathbb{R}$ s.t. $H \subseteq \{x : a^T x > b\}$, $M \subseteq \{x : a^T x \le b\}$

Pattern Classification via LP

Lemma 1:

$H = \{H^1, H^2, \ldots, H^h\} \subseteq \mathbb{R}^n$

$M = \{M^1, M^2, \ldots, M^m\} \subseteq \mathbb{R}^n$

H and M are linearly separable

$\Leftrightarrow \exists\ a \in \mathbb{R}^n,\ b \in \mathbb{R}$ s.t. $a^T H^i - b \geq +1\ \forall i = 1..h$, $a^T M^j - b \leq -1\ \forall j = 1..m$

Proof

$\Leftarrow$ TRIVIAL

$\Rightarrow$

Pattern Classification via LP

Lemma 1:

H and M are linearly separable $\Leftrightarrow \exists \; a \in \mathbb{R}^n, \; b \in \mathbb{R}$ s.t.

$$a^T H^i - b \geq +1 \quad \forall i = 1..h$$
$$a^T M^j - b \leq -1 \quad \forall j = 1..m$$

$\Rightarrow$ IS A LITTLE MORE COMPLICATED

Proof

$$\Rightarrow \exists \; c \in \mathbb{R}^n, \; b \in \mathbb{R} : \quad c^T x > b \quad \forall x \in H$$
$$\leq b \quad \forall x \in M$$

LET $\rho \doteq \min_{x \in H} c^T x - \max_{x \in M} c^T x > 0$

Pattern Classification via LP

Proof continued

$$\Rightarrow \text{LET } \rho \doteq \min_{x \in H} c^T x - \max_{x \in M} c^T x > 0$$

$$a \doteq \frac{2}{\rho} c \qquad b \doteq \frac{1}{\rho}\left[\min_{x \in H} c^T x + \max_{x \in M} c^T x\right]$$

$$\text{Now, } \min_{x \in H} a^T x = \min \frac{2}{\rho} c^T x = \min \frac{1}{\rho}\left[c^T x + c^T x\right]$$

$$= \frac{1}{\rho} \min \left[c^T x + \max_{x \in M} c^T x + \rho\right]$$

$$= \frac{1}{\rho}\left[\min c^T x + \max c^T x + \rho\right] = b + 1$$

$$\min_{x \in H} a^T x = b + 1$$

Pattern Classification via LP

Proof continued

$$\Rightarrow \text{LET } \rho \doteq \min_{x \in H} c^T x - \max_{x \in M} c^T x > 0$$

$$\text{LET } a \doteq \frac{2}{\rho} c \in \mathbb{R}^n, \quad b \doteq \frac{1}{\rho}\left[\min_{x \in H} c^T x + \max_{x \in M} c^T x\right]$$

Similarly, $\max_{x \in M} a^T x = \max_{x \in M} \frac{2}{\rho} c^T x = \frac{1}{\rho}\left[\underbrace{\max_{x \in M} c^T x}_{\min_{x \in H} c^T x - \rho} + \max_{x \in M} c^T x\right]$

$$= \frac{1}{\rho}\left[-\rho + \underbrace{\min_{x \in H} c^T x + \max_{x \in M} c^T x}_{\rho b}\right] = -1 + b$$

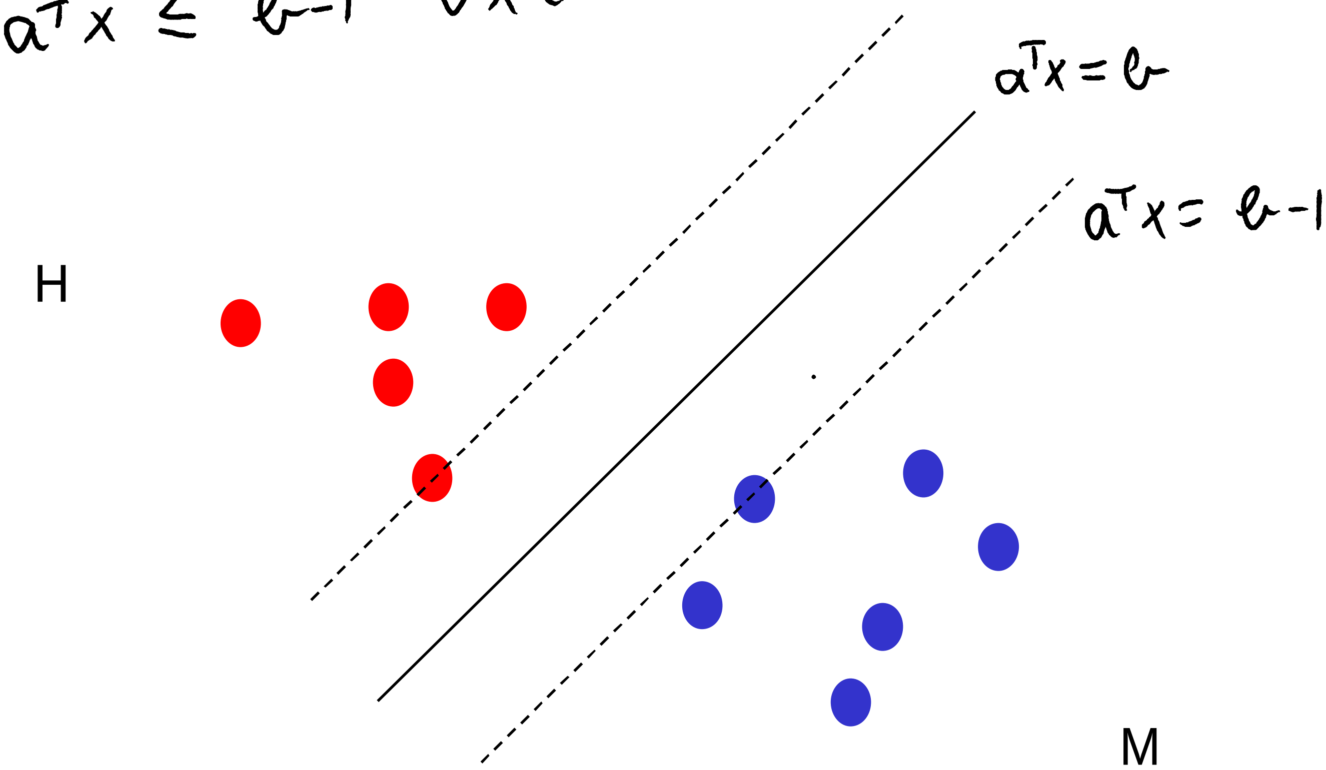
$$\Rightarrow \max_{x \in M} a^T x = b - 1$$

Pattern Classification via LP

Proof continued

$$\Rightarrow \quad a^T x \geq b+1 \quad \forall x \in H \qquad \square$$

$$a^T x \leq b-1 \quad \forall x \in M$$



Pattern Classification via LP

We will see that the following linear problem solves Problem 1 & 2:

[Mansgarian 1995]

GIVEN SETS $H = \{H^1, H^2, \ldots, H^h\} \subseteq \mathbb{R}^n$

$M = \{M^1, M^2, \ldots, M^m\} \subseteq \mathbb{R}^n$

FIND $y \in \mathbb{R}^h$, $z \in \mathbb{R}^m$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}$ SUCH THAT

$$\min \frac{1}{h}\left[y_1 + y_2 + \ldots + y_h\right] + \frac{1}{m}\left[z_1 + z_2 + \ldots + z_m\right]$$

$y \in \mathbb{R}^h$, $z \in \mathbb{R}^m$, $a \in \mathbb{R}^n$, $b \in \mathbb{R}$

s.t.

$$y_i \geq -a^T H^i + b + 1 \quad \forall i = 1 \ldots h$$

$$z_j \geq a^T M^j - b + 1 \quad \forall j = 1 \ldots m$$

$$y_i \geq 0 \quad \forall i = 1 \ldots h$$

$$z_j \geq 0 \quad \forall j = 1 \ldots m$$

Pattern Classification via LP

FIND $y \in \mathbb{R}^k, z \in \mathbb{R}^m, a \in \mathbb{R}^n, b \in \mathbb{R}$ SUCH THAT

$$
\left.
\begin{aligned}
&\min \frac{1}{k}\left[y_1 + y_2 + \dots + y_k\right] + \frac{1}{m}\left[z_1 + z_2 + \dots + z_m\right] \\
&y \in \mathbb{R}^k \quad \text{s.t.} \quad y_i \geq -a^T H^i + b + 1 \quad \forall i = 1 \dots k \\
&z \in \mathbb{R}^m \qquad\quad z_j \geq a^T M^j - b + 1 \quad \forall j = 1 \dots m \\
&a \in \mathbb{R}^n \qquad\quad y_i \geq 0 \quad \forall i = 1 \dots k \\
&b \in \mathbb{R} \qquad\quad\;\; z_j \geq 0 \quad \forall j = 1 \dots m
\end{aligned}
\right\} (LP)
$$

Theorem 1

H and M are linearly separable iff the optimal value of LP is 0.

Theorem 2

H and M are linearly separable

y^*, z^*, a^*, b^* is an optimal solution of (LP)

⇒ $f(x) = a^{*T}x + b^*$ is a separating hyperplane

Pattern Classification via LP

FIND $y \in \mathbb{R}^k, z \in \mathbb{R}^m, a \in \mathbb{R}^n, b \in \mathbb{R}$ SUCH THAT

$$\min_{\substack{y \in \mathbb{R}^k \\ z \in \mathbb{R}^m \\ a \in \mathbb{R}^n \\ b \in \mathbb{R}}} \frac{1}{k}\left[y_1 + y_2 + \ldots + y_k\right] + \frac{1}{m}\left[z_1 + z_2 + \ldots + z_m\right]$$

$$\text{s.t.} \quad y_i \geq -a^T H^i + b + 1 \quad \forall i = 1 \ldots k$$

$$z_j \geq a^T M^j - b + 1 \quad \forall j = 1 \ldots m$$

$$y_i \geq 0 \quad \forall i = 1 \ldots k$$

$$z_j \geq 0 \quad \forall j = 1 \ldots m$$

(LP)

Proof of Theorems 1 and 2

The optimal value of (LP) is 0 $\Longleftrightarrow$

$$y^* = 0$$

$$z^* = 0$$

$$a^T H^i \geq b^* + 1 \quad \forall i = 1 \ldots k$$

$$a^T M^i \leq b^* - 1 \quad \forall j = 1 \ldots m$$

Pattern Classification via LP

Application: Breast Cancer Diagnosis

Used at the University of Wisconsin Hospital

[Mangasarian et al 1995]

1. Fluid sample from breast.
2. Placed on a glass and stained the highlight the nuclei of cells
3. Image is taken
4. 30D features: Area, Radius, perimeter, etc

Goal: Classification between benign lumps and malignant lumps

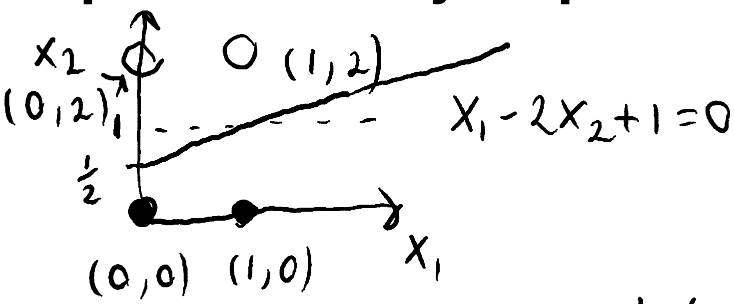
Results: 97.5% accuracy

Pattern Classification via LP

Example 1: Linearly Separable Case

$H = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$ $k=2$

$M = \left\{ \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ $m=2$



LP: $\min_{a_1, a_2, b, y_1, y_2, z_1, z_2} \frac{1}{2}(y_1 + y_2) + \frac{1}{2}(z_1 + z_2)$

S.T.

$y_1 \geq -[a_1, a_2] \begin{bmatrix} H_1^1 \\ H_2^1 \end{bmatrix} + b + 1 = b + 1$ ← $\begin{pmatrix} 0 \\ 0 \end{pmatrix}$

$y_2 \geq -[a_1, a_2] \begin{bmatrix} H_1^2 \\ H_2^2 \end{bmatrix} + b + 1 = -a_1 + b + 1$ ← $\begin{pmatrix} 1 \\ 0 \end{pmatrix}$

$z_1 \geq [a_1, a_2] \begin{bmatrix} M_1^1 \\ M_2^1 \end{bmatrix} - b + 1 = 2a_2 - b + 1$ ← $\begin{bmatrix} 0 \\ 2 \end{bmatrix}$

$z_2 \geq [a_1, a_2] \begin{bmatrix} M_1^2 \\ M_2^2 \end{bmatrix} - b + 1 = a_1 + 2a_2 - b + 1$ ← $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$

$y_1, y_2, z_1, z_2 \geq 0$

(LP)

AN OPTIMAL SOLUTION

$y_1 = y_2 = z_1 = z_2 = 0$

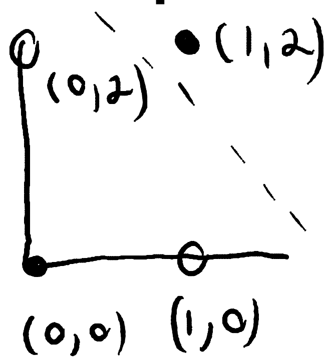
$a^T = [1, -2]$ $b = -1$

$a^T = [0, -1]$ $b = -1$

$0x_1 - 1x_2 + 1 = 0$

Pattern Classification via LP

Example 2: Linearly nonseparable case $H = \left\{ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$ $M = \left\{ \begin{pmatrix} 0 \\ 2 \end{pmatrix}, \begin{pmatrix} 1 \\ 0 \end{pmatrix} \right\}$



$$\min_{Y_1, Y_2, Z_1, Z_2, a, b} \ \frac{1}{2}(Y_1 + Y_2) + \frac{1}{2}(Z_1 + Z_2)$$

S.T.

$$Y_1 \geq -[a_1 \ a_2] \begin{bmatrix} H_1^1 \\ H_2^1 \end{bmatrix} + b + 1 = b + 1 \qquad \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

$$Y_2 \geq -[a_1 \ a_2] \begin{bmatrix} H_1^2 \\ H_2^2 \end{bmatrix} + b + 1 = -a_1 - 2a_2 + b + 1 \qquad \begin{pmatrix} 1 \\ 2 \end{pmatrix}$$

$$Z_1 \geq [a_1 \ a_2] \begin{bmatrix} M_1^1 \\ M_2^1 \end{bmatrix} - b + 1 = 2a_2 - b + 1 \qquad \begin{pmatrix} 0 \\ 2 \end{pmatrix}$$

$$Z_2 \geq [a_1, a_2] \begin{bmatrix} M_1^2 \\ M_2^2 \end{bmatrix} - b + 1 = a_1 - b + 1 \qquad \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$Y_1, Y_2, Z_1, Z_2 \geq 0$$

AN OPTIMAL SOLUTION:

$Y_1 = 4$, $Y_2 = 0$, $Z_1 = 0$, $Z_2 = 0$

$b = 3$, $a = [2, 1]$ $\Rightarrow$ $2X_1 + X_2 - 3 = 0$ IS THE PROPOSED NONSEPARABLE HYPER PLANE

Linear Programs

- Standard from, Canonical form, Inequality form
- Transforming LPS
  - Pivot transformation

Linear Programs

Inequality form of LPs using matrix notation:

$$\text{MIN [OR MAX]} \quad c^T x \qquad c \in \mathbb{R}^n$$

$$\text{s.t.} \quad Ax \le b \qquad A \in \mathbb{R}^{m \times n} \quad b \in \mathbb{R}^m$$

$$l \le x \le u \qquad l \in \mathbb{R}^n \quad u \in \mathbb{R}^n$$

$$x \in \mathbb{R}^n$$

Standard form of LPs:

$$\text{MIN} \quad c^T x \quad c \in \mathbb{R}^n$$

$$\text{s.t.} \quad Ax = b$$

$$x \ge 0 \qquad b \ge 0$$

Theorem: Any LP can be rewritten to an equivalent standard LP

Transforming LPs

Theorem: Any LP can be rewritten to an equivalent standard LP

- Getting rid of inequalities (except variable bounds)

$$x_1 + x_2 \leq 4 \Rightarrow x_1 + x_2 + x_3 = 4$$
$$x_3 \geq 0$$

- Getting rid of equalities

$$x_1 + x_2 = 4 \Leftrightarrow x_1 + x_2 \geq 4$$
$$\leq 4$$

Transforming LPs

- Getting rid of negative variables

  $x \in \mathbb{R} \quad x = u - v \quad u \in \mathbb{R}_{+}, \; v \in \mathbb{R}_{+}$

- Getting rid of bounded variables

  $x \in [2,5] \quad 2 \le x, \; x \le 5$

- Max to Min

  $\mathrm{MAX}\ c^T x = -\mathrm{MIN}(-c^T x)$

- Negative b_i

  $a_i^T x = b_i \iff -a_i^T x = -b_i$

From Inequality Form to Standard Form

Inequality form

max 2x+3y s.t.

- x + y · 4
- 2x + 5y · 12
- x + 2y · 5
- x, y ¸ 0

if std fm has n vars, m eqns,

then ineq form has

n–m vars and m+(n–m)=n ineqs

(here m = 3, n = 5)

Standard form

max 2x+3y s.t.

- x + y + u = 4
- 2x + 5y + v = 12
- x + 2y + w = 5
- x, y, u, v, w ¸ 0

SLACK VARIABLES