Convex Optimization
CMU-10725

2. Linear Programs

Barnabás Póczos & Ryan Tibshirani

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

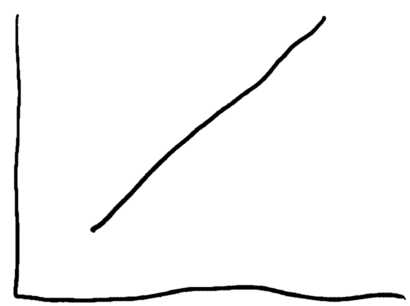
- More and more complicated optimization problems
- Definition of LP

Simplest Optimization Problems

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

- Constant function $f(x) = c \quad x \in \mathbb{R}^n$
- 1-dim linear function $f(x) = ax + b$

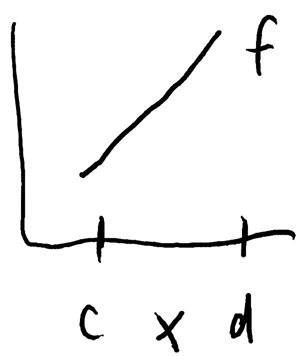
$$\max_x f(x) = \infty \qquad \min_x f(x) = -\infty$$



- 1-dim linear function with bound constraints

$x \geq c \Rightarrow \arg\min f(x) = c$ if $a > 0$

$x \leq d \Rightarrow \arg\max f(x) = d$ if $a > 0$



Linear Programs

- n-dim linear function with m linear constraints

Inequality form:

Cost function: MIN [OR MAX] $c_1 x_1 + c_2 x_2 + \dots + c_n x_n$

Constraints:

$$\text{S.T.} \quad a_{11} x_1 + a_{12} x_2 + \dots + a_{1n} x_n \le b_1$$
$$\vdots$$
$$a_{m1} x_1 + a_{m2} x_2 + \dots + a_{mn} x_n \le b_m$$

Bounds:

$$l_1 \le x_1 \le u_1$$
$$\vdots$$
$$l_m \le x_m \le u_m$$

l_i CAN BE $-\infty$
u_i CAN BE ∞

Linear Programs

Inequality form using matrix notation:

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

Constraints: $\text{S.T} \quad Ax \leq b \qquad A \in \mathbb{R}^{m \times n} \qquad b \in \mathbb{R}^m$

Bounds: $l \leq x \leq u \qquad l \in \mathbb{R}^n \quad u \in \mathbb{R}^n$

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

Example:

$$\begin{aligned} \text{MIN} \quad & -2x_1 - x_2 \\ \text{S.T} \quad & x_1 + x_2 \leq 5 \\ & 2x_1 + 3x_2 \leq 12 \\ & x_1 \leq 4 \\ & x_1 \geq 0, \; x_2 \geq 0 \end{aligned}$$

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

$$A = \begin{bmatrix} 1, & 1 \\ 2, & 3 \end{bmatrix} \qquad b = \begin{bmatrix} 5 \\ 12 \end{bmatrix}$$

$$l = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \qquad u = \begin{bmatrix} 4 \\ \infty \end{bmatrix}$$

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	

$X_1 \geq 0 \quad X_2 \geq 0, \quad X_3 \geq 0, \quad X_4 \geq 0, \quad X_5 \geq 0, \quad X_6 \geq 0$

$\text{MAX PROFIT} = 12X_1 + 20X_2 + 18X_3 + 40X_4$

$\text{S.T.} \quad 4X_1 + 9X_2 + 7X_3 + 10X_4 + X_5 \leq 6000$

$X_1 + X_2 + 3X_3 + 40X_4 + X_6 \leq 4000$

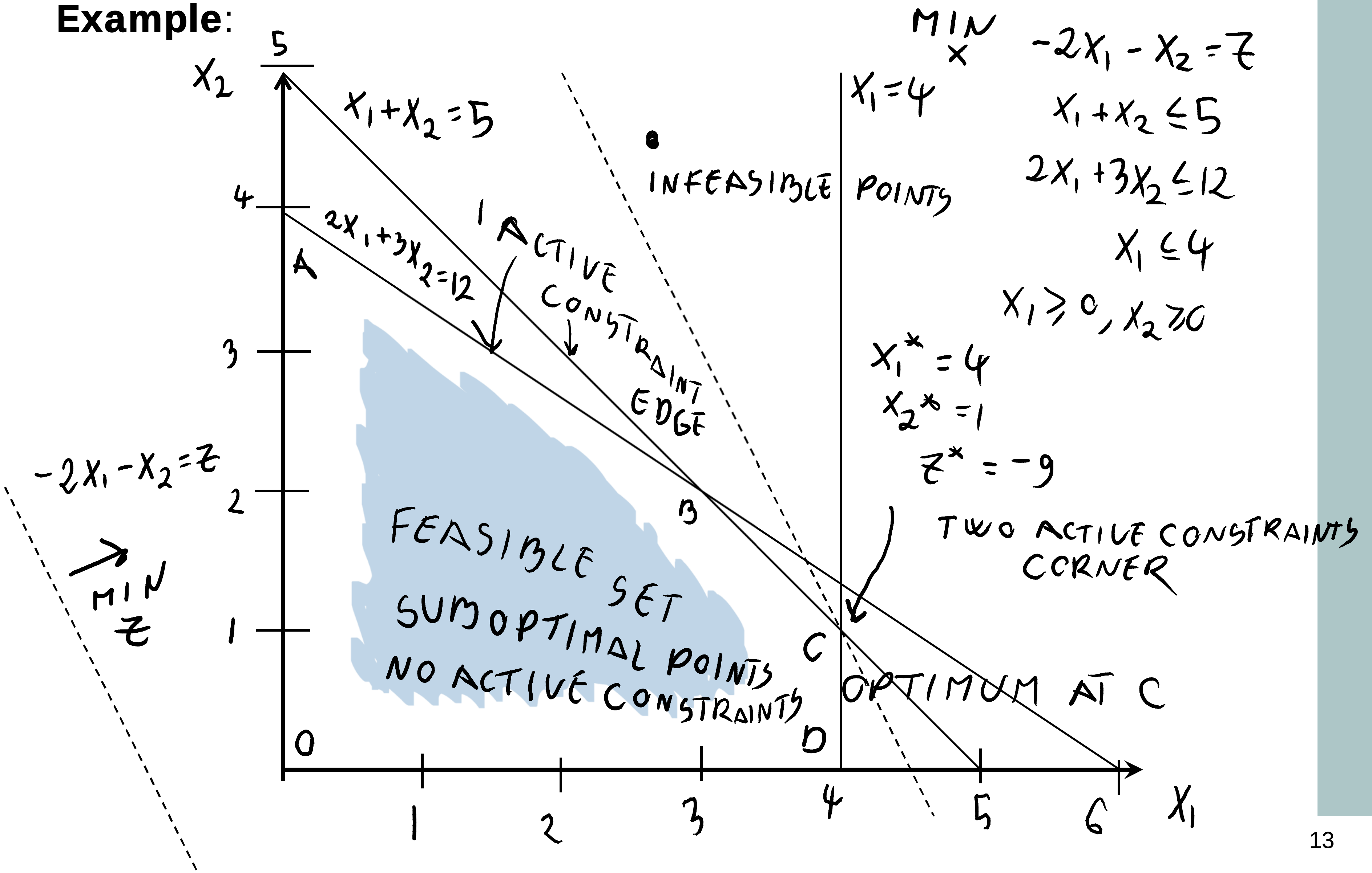
Why is it called Linear Programing???

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

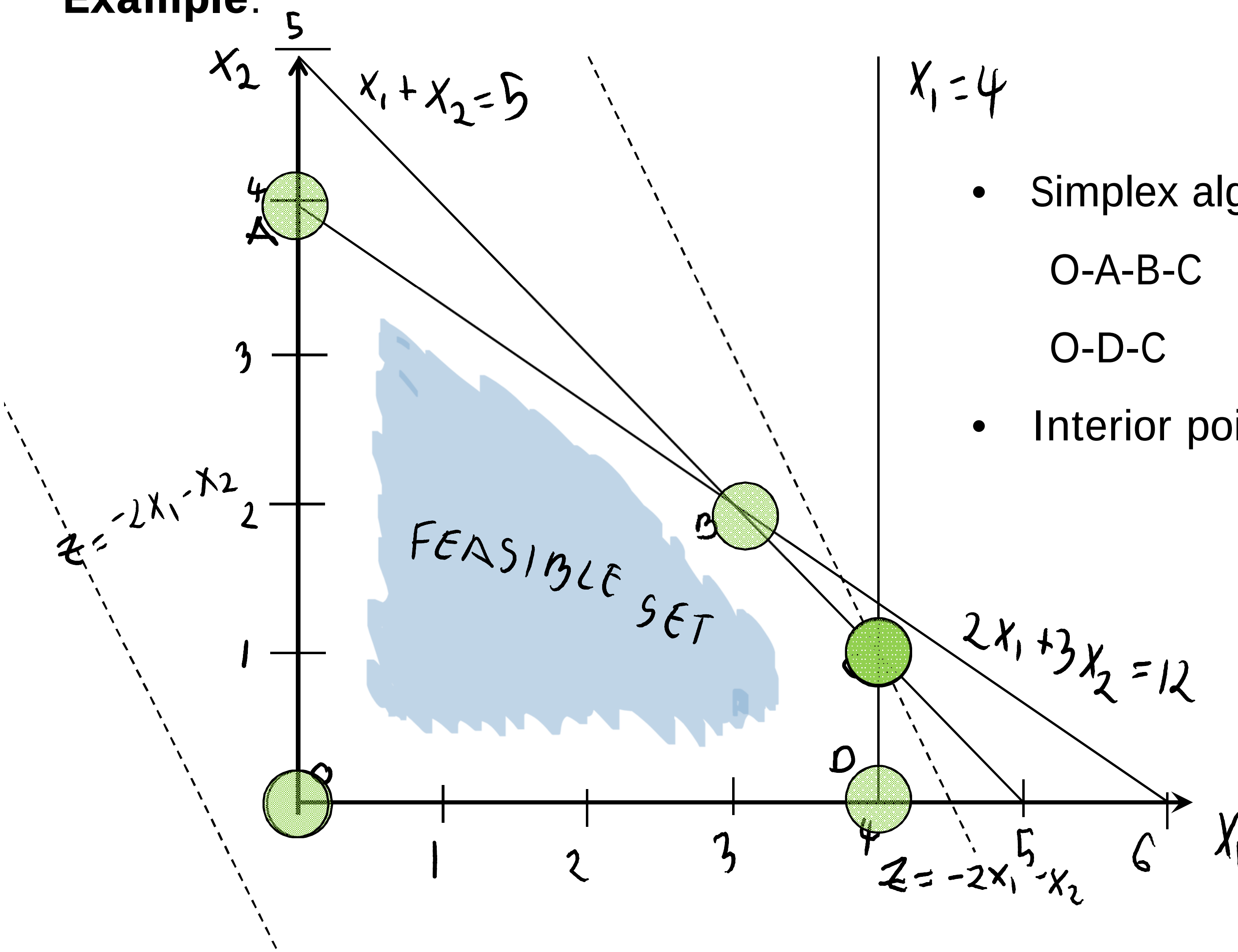
Sketching Linear Programs

Example:



Simplex Algorithm

Example:



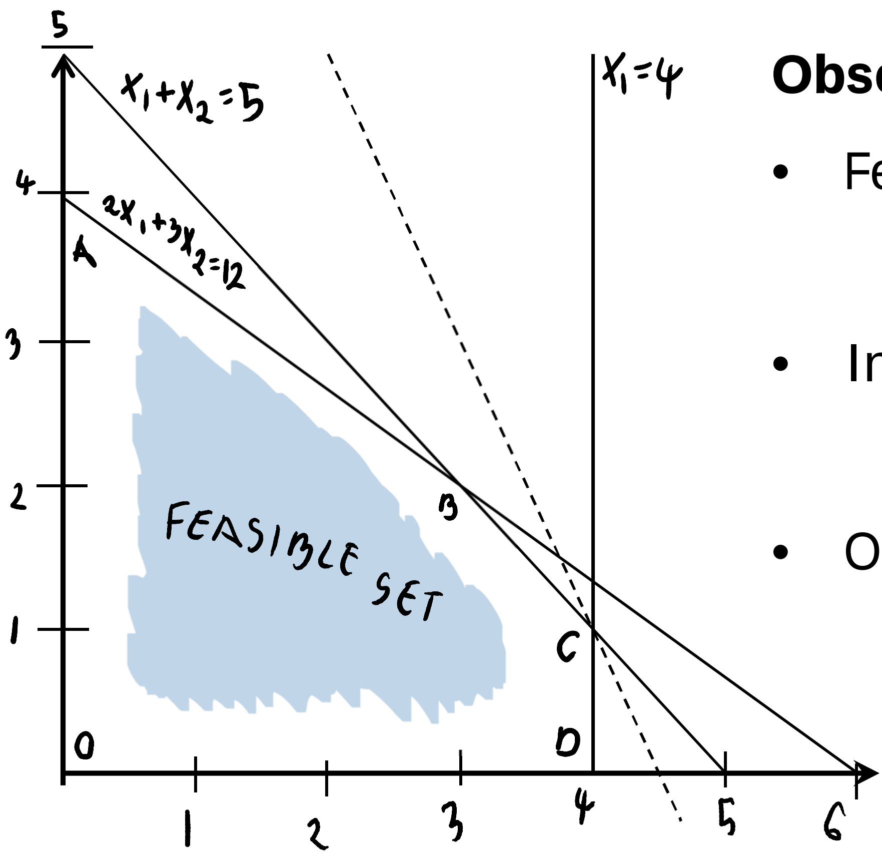
- Simplex algorithm:

  O-A-B-C

  O-D-C

- Interior point methods

Linear Program



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)

$$\begin{aligned} &\text{MIN [OR MAX]} \; c^T x && c \in \mathbb{R}^n \\ &\text{s.t.} \; Ax \le b && A \in \mathbb{R}^{m \times n} \\ &\quad l \le x \le u && l, u, x \in \mathbb{R}^n \end{aligned}$$

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

$l \leq x \leq u$ $\quad l, u, x \in \mathbb{R}^n$

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:

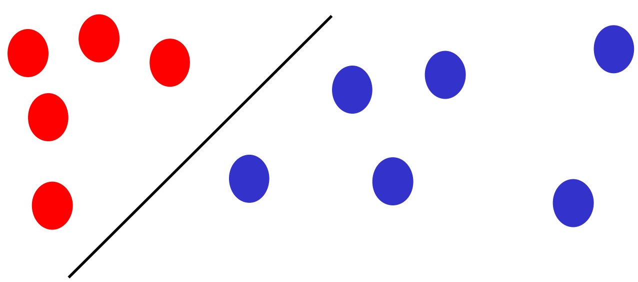
GIVEN $H = \{H^1, H^2, \ldots, H^h\} \subseteq \mathbb{R}^n$ DISJOINT SETS

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

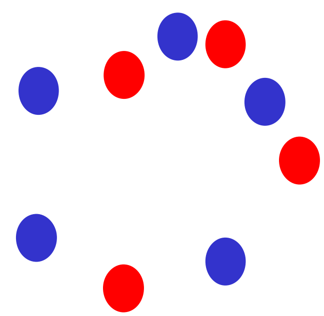
Problem 1: Determine whether H and M are linearly separable

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

Linearly separable sets:

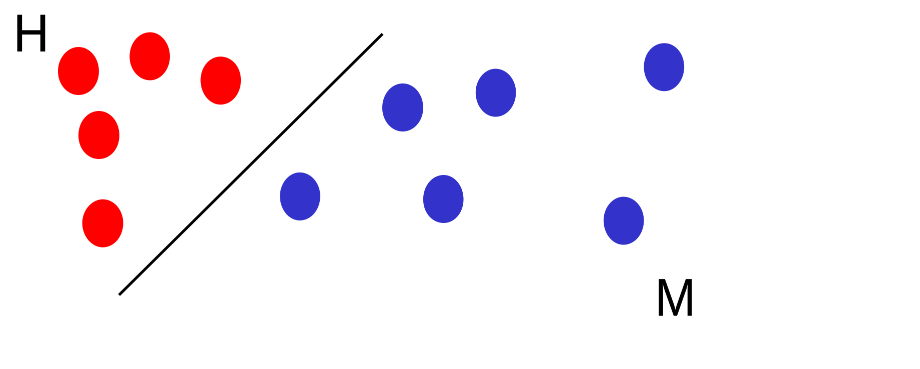


Linearly not separable sets:

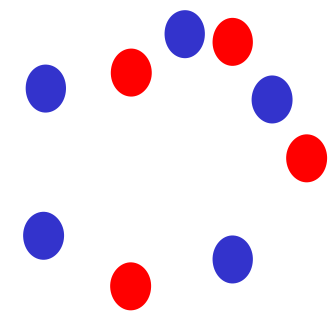


Pattern Classification via LP

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. $H \subseteq \{x \in \mathbb{R}^n : a^T x > b\}$, $M \subseteq \{x \in \mathbb{R}^n : a^T x \leq b\}$

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

$$\Leftrightarrow \exists \begin{array}{l} a \in \mathbb{R}^n \\ b \in \mathbb{R} \end{array} \text{ s.t. } \begin{array}{l} a^T H^i - b \geq +1 \quad \forall i = 1 \ldots h \\ a^T M^j - b \leq -1 \quad \forall j = 1 \ldots m \end{array}$$

Proof

$\Leftarrow$ TRIVIAL

$\Rightarrow$ SINCE $f(x) = a^T x - b$ SEPARATES H AND M

THEREFORE, $\{x : a^T x = b\}$ IS A SEPARATING HYPERPLANE

$\Rightarrow$ H AND M ARE LINEARLY SEPARABLE.

Pattern Classification via LP

Lemma 1:

H and M are linearly separable

$$\iff \exists \; \begin{matrix} a \in \mathbb{R}^n \\ b \in \mathbb{R} \end{matrix} \text{ s.t. } \begin{matrix} a^T H^i - b \geq +1 \quad \forall i = 1..h \\ a^T M^j - b \leq -1 \quad \forall j = 1..m \end{matrix}$$

$\Rightarrow$ IS A LITTLE MORE COMPLICATED

Proof

$$\Rightarrow \exists\, c \in \mathbb{R}^n, b \in \mathbb{R} : \begin{matrix} c^T x > b & \forall x \in H \\ \leq b & \forall x \in M \end{matrix}$$

$$\Rightarrow \min_{x \in H} c^T x > b \geq \max_{x \in M} c^T x$$

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

Pattern Classification via LP

Proof continued

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

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

$$\text{NOW, } \min_{x\in H} a^T x = \min_{x\in H} \frac{2}{\rho} c^T x = \frac{1}{\rho}\Big[\min_{x\in H} c^T x + \underbrace{\min_{x\in H} c^T x}_{\max_{x\in M} c^T x + \rho \ [\text{BY DEF } \rho]}\Big]$$

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

$$= b + 1 \quad [\text{BY DEF OF } b]$$

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

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

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

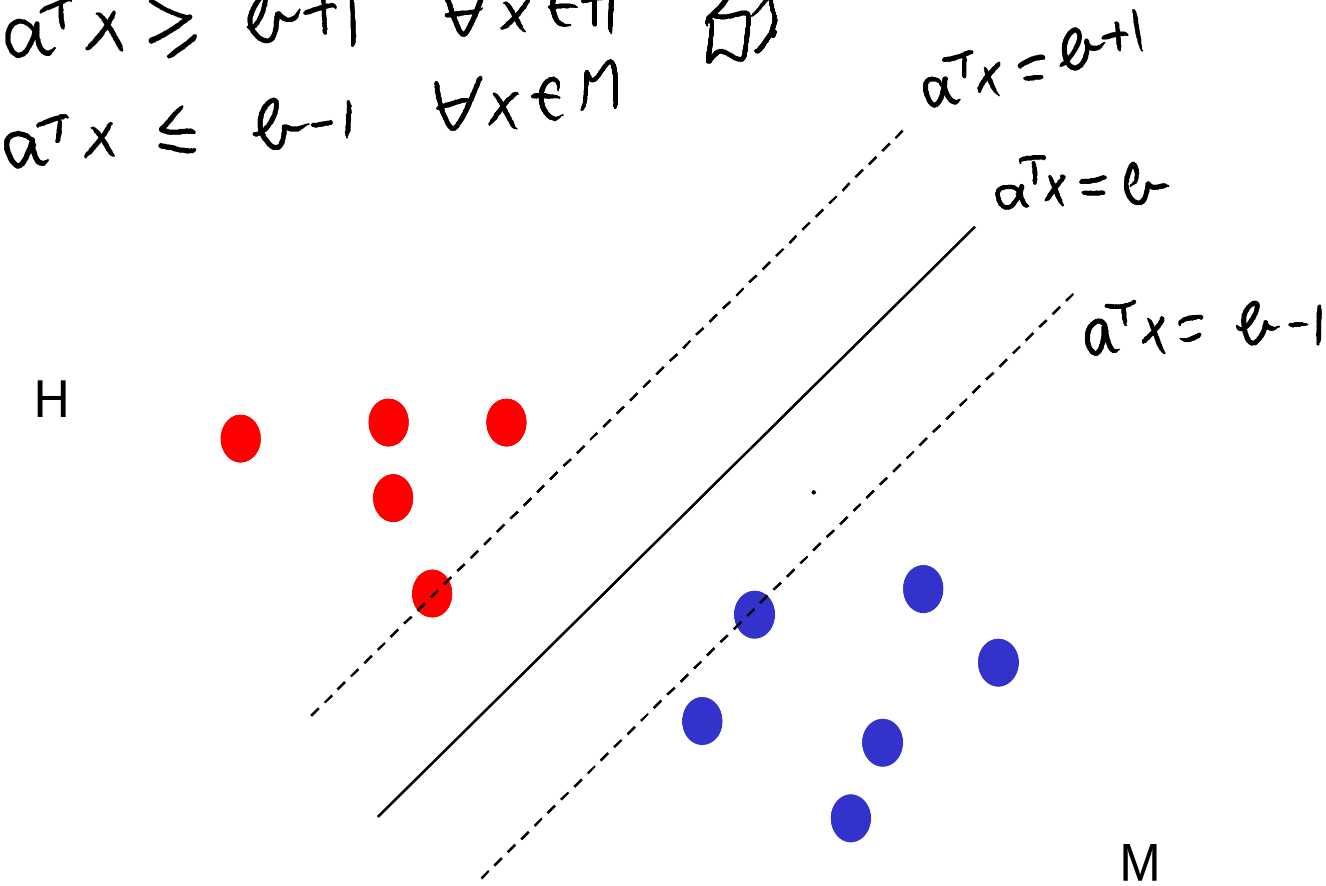
$$\Rightarrow \max_{x \in M} a^T x = b - 1$$

Pattern Classification via LP

Proof continued

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

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



Pattern Classification via LP

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

GIVEN SETS $H = \{H^1, H^2, \dots, H^h\} \subseteq \mathbb{R}^n$ [Mansgarian 1995]

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

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

$$\text{MIN } \frac{1}{h}\left[y_1 + y_2 + \dots + y_h\right] + \frac{1}{m}\left[z_1 + z_2 + \dots + z_m\right]$$

$y \in \mathbb{R}^h$ S.T. $y_i \geq -a^T H^i + b + 1 \quad \forall i = 1 \dots h$

$z \in \mathbb{R}^m$ $\quad z_j \geq a^T M^j - b + 1 \quad \forall j = 1 \dots m$

$a \in \mathbb{R}^n$ $\quad y_i \geq 0 \quad \forall i = 1 \dots h$

$b \in \mathbb{R}$

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

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

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

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)

$\Rightarrow$ $f(x)=a^{*T}x+b^*$ 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

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

Proof of Theorems 1 and 2

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

$$
\begin{array}{ll}
(1) & y^* = 0 \\
(2) & z^* = 0 \\
(3) & a^{*T} H^i \geq b^* + 1 \quad \forall i = 1 \ldots k \\
(4) & a^{*T} M^j \leq b^* - 1 \quad \forall j = 1 \ldots m
\end{array}
$$

$\Longleftrightarrow$ H AND M ARE LINEARLY SEPARABLE

$\{x : a^{*T} x - b^* = 0\}$ IS A SEPARATING HYPERPLANE $\square$

Pattern Classification via LP

Application: Breast Cancer Diagnosis

Used at the University of Wisconsin Hospital

[Mangasarian et al 1995]

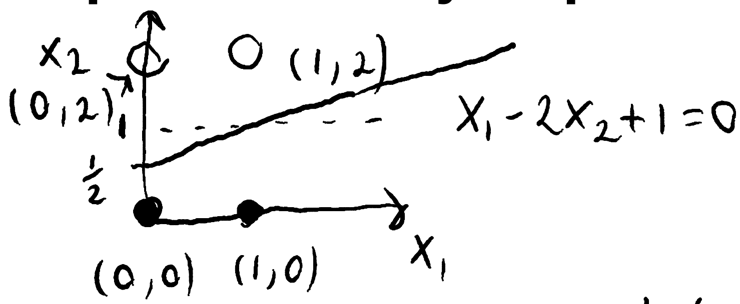
1. Fluid sample from breast.
2. Placed on a glass and stained the highlight the nuclei of cells
3. Image is taken
4. 30D features: Area, Radius, perimeter, etc

Goal: Classification between benign lumps and malignant lumps

Results: 97.5% accuracy

Pattern Classification via LP

Example 1: Linearly Separable Case



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

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

LP: $\min\limits_{a_1, a_2, b, Y_1, Y_2, Z_1, Z_2} \ \frac{1}{2}(Y_1 + Y_2) + \frac{1}{2}(Z_1 + Z_2)$

S.T.

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

$$Y_2 \geq -[a_1, a_2] \begin{bmatrix} H_1^2 \\ H_2^2 \end{bmatrix} + b + 1 = -a_1 + b + 1 \qquad \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

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

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

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

(LP)

AN OPTIMAL SOLUTION

$Y_1 = Y_2 = Z_1 = Z_2 = 0$

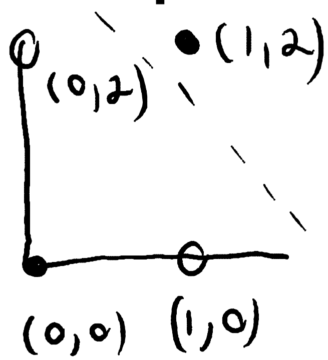
$a^T = [1, -2] \quad b = -1$

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

$0x_1 - 1x_2 + 1 = 0$

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.

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

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

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

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

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

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:

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

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

$$\ell \leq x \leq u \qquad \ell \in \mathbb{R}^n \quad u \in \mathbb{R}^n$$

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

Standard form of LPs:

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

$$\text{S.T} \quad Ax = b$$

$$x \geq 0 \qquad x \in \mathbb{R}^n_{+}$$

$$b \in \mathbb{R}^m_{+}$$

SOMETIMES $b \geq 0$ IS NOT REQUIRED

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)

$$x_1 + x_2 \leq 4 \implies x_1 + x_2 + x_3 = 4$$

$$x_3 \geq 0$$

SLACK VARIABLE

- Getting rid of equalities

$$x_1 + 2x_2 = 4 \implies \begin{array}{l} x_1 + 2x_2 \leq 4 \\ x_1 + 2x_2 \geq 4 \end{array}$$

Transforming LPs

- Getting rid of negative variables

$$x \in \mathbb{R} \Rightarrow x = u - v \qquad \begin{array}{l} u \geq 0 \\ v \geq 0 \end{array}$$

- Getting rid of bounded variables

$$x \in [2, 5] \Rightarrow \begin{cases} 2 \leq x \\ x \leq 5 \end{cases}$$

- Max to Min

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

- Negative b_i

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