Lecture 12: Linear Programming and Duality Theorems

Rafael Oliveira

University of Waterloo
Cheriton School of Computer Science

rafael.oliveira.teaching@gmail.com

June 17, 2021

Overview

- Part I
  - Why Linear Programming?
  - Structural Results on Linear Programming
  - Duality Theory
- Part II
  - Game Theory
  - Learning Theory - Boosting
- Conclusion
- Acknowledgements

Mathematical Programming

Mathematical Programming deals with problems of the form

Mathematical Programming

Mathematical Programming deals with problems of the form

$$\begin{array}{rl} \text{minimize} & \underbrace{f(x)}_{\text{objective function}} \\ \text{subject to} & \left.\begin{array}{l} g_1(x) \leq 0 \\ \quad\vdots \\ g_m(x) \leq 0 \end{array}\right\} \text{Constraints} \\ & x \in \mathbb{R}^n \end{array}$$

Mathematical Programming

Mathematical Programming deals with problems of the form

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_1(x) \leq 0 \\ & \quad \vdots \\ & g_m(x) \leq 0 \\ & x \in \mathbb{R}^n \end{aligned}$$

- Very general family of problems.

For instance: NP-hard when g_i's are quadratic polynomials!

Question: how much harder can it get?
(much harder!)

Mathematical Programming

Mathematical Programming deals with problems of the form

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_1(x) \leq 0 \\ & \quad \vdots \\ & g_m(x) \leq 0 \\ & x \in \mathbb{R}^n \end{aligned}$$

- Very general family of problems.
- Special case is when all functions $f, g_1, \ldots, g_m$ are *linear* functions (called *Linear Programming* - LP for short)

Mathematical Programming

Mathematical Programming deals with problems of the form

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_1(x) \leq 0 \\ & \quad \vdots \\ & g_m(x) \leq 0 \\ & x \in \mathbb{R}^n \end{aligned}$$

- Very general family of problems.
- Special case is when all functions $f, g_1, \ldots, g_m$ are *linear* functions (called *Linear Programming* - LP for short)
- Traces of idea of LP in works of Fourier [Fourier 1823, Fourier 1824]

Mathematical Programming

Mathematical Programming deals with problems of the form

$$\begin{aligned} \text{minimize} \quad & f(x) \\ \text{subject to} \quad & g_1(x) \le 0 \\ & \quad \vdots \\ & g_m(x) \le 0 \\ & x \in \mathbb{R}^n \end{aligned}$$

- Very general family of problems.
- Special case is when all functions $f, g_1, \ldots, g_m$ are *linear* functions (called *Linear Programming* - LP for short)
- Traces of idea of LP in works of Fourier [Fourier 1823, Fourier 1824]
- Formally studied & importance of LP recognized in 1940's by Dantzig, Kantorovich, Koopmans and von Neumann.

What is a Linear Program?

A linear function $f : \mathbb{R}^n \to \mathbb{R}$ is given by

$$f(x) = \underbrace{c_1}_{\in \mathbb{R}} \cdot x_1 + \ldots + \underbrace{c_n}_{\in \mathbb{R}} \cdot x_n = c^T x$$

$$c = \begin{pmatrix} c_1 \\ \vdots \\ c_n \end{pmatrix} \qquad c^T = \begin{pmatrix} c_1 & c_2 & \cdots & c_n \end{pmatrix}$$

What is a Linear Program?

A linear function $f : \mathbb{R}^n \to \mathbb{R}$ is given by

$$f(x) = c_1 \cdot x_1 + \ldots + c_n \cdot x_n = c^T x$$

Linear Programming deals with problems of the form

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & A_1^T x \leq 0 \\ & \vdots \\ & A_m^T x \leq 0 \\ & x \in \mathbb{R}^n \end{aligned}$$

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

What is a Linear Program?

A linear function $f : \mathbb{R}^n \to \mathbb{R}$ is given by

$$f(x) = c_1 \cdot x_1 + \ldots + c_n \cdot x_n = c^T x$$

Linear Programming deals with problems of the form

$$\begin{array}{rl} \text{minimize} & c^T x \\ \text{subject to} & Ax \leq 0 \\ & x \in \mathbb{R}^n \end{array}$$

$$A \in \mathbb{R}^{m \times n}$$

$$\begin{pmatrix} A_1^T \\ A_2^T \\ \vdots \\ A_n^T \end{pmatrix}$$

What is a Linear Program?

A linear function $f : \mathbb{R}^n \to \mathbb{R}$ is given by

$$f(x) = c_1 \cdot x_1 + \ldots + c_n \cdot x_n = c^T x$$

Linear Programming deals with problems of the form

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax \leq 0 \\ & x \in \mathbb{R}^n \end{aligned}$$

We can *always* represent LPs in *standard form*:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

$A \in \mathbb{R}^{m \times n}$
$b \in \mathbb{R}^m$

Practice problem: show that we can always represent a LP in standard form.

Why should I care?

- Linear Programs appear everywhere in life: many problems of interest (resource allocation problems) can be modelled as linear program!

Why should I care?

- Linear Programs appear everywhere in life: many problems of interest (resource allocation problems) can be modelled as linear program!
- Stock portfolio optimization:

Why should I care?

- Linear Programs appear everywhere in life: many problems of interest (resource allocation problems) can be modelled as linear program!
- Stock portfolio optimization:
  - n companies, stock of company i costs $c_i \in \mathbb{R}$
  - company i has expected profit $p_i \in \mathbb{R}$
  - our budget is $B \in \mathbb{R}$

Why should I care?

- Linear Programs appear everywhere in life: many problems of interest (resource allocation problems) can be modelled as linear program!
- Stock portfolio optimization:
  - n companies, stock of company i costs $c_i \in \mathbb{R}$
  - company i has expected profit $p_i \in \mathbb{R}$
  - our budget is $B \in \mathbb{R}$

total profit

$$\begin{array}{rl} \text{maximize} & \boxed{p_1 \cdot x_1 + \cdots + p_n \cdot x_n} \\ \text{subject to} & c_1 \cdot x_1 + \cdots + c_n \cdot x_n \leq B \\ & x \geq 0 \end{array}$$

cannot spend more than B to create our portfolio

Variables: x_i amount of stock i that you want to have

$x_i \cdot p_i$ profit from stock i

Why should I care?

- Linear Programs appear everywhere in life: many problems of interest (resource allocation problems) can be modelled as linear program!
- Stock portfolio optimization:
  - n companies, stock of company i costs $c_i \in \mathbb{R}$
  - company i has expected profit $p_i \in \mathbb{R}$
  - our budget is $B \in \mathbb{R}$

$$\begin{aligned} \text{maximize} \quad & p_1 \cdot x_1 + \cdots + p_n \cdot x_n \\ \text{subject to} \quad & c_1 \cdot x_1 + \cdots + c_n \cdot x_n \leq B \\ & x \geq 0 \end{aligned}$$

- Other problems, such as *data fitting, linear classification* can be modelled as linear programs.

Important Questions

$$
\begin{aligned}
\text{minimize} \quad & c^T x \\
\text{subject to} \quad & Ax = b \\
& x \geq 0
\end{aligned}
$$

Important Questions

$$
\begin{array}{rl}
\text{minimize} & c^T x \\
\text{subject to} & \boxed{\begin{array}{l} Ax = b \\ x \geq 0 \end{array}}
\end{array}
$$

1. When is a Linear Program *feasible*?
   - Is there a solution to the constraints at all?

Important Questions

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

1. When is a Linear Program *feasible*?
   - Is there a solution to the constraints at all?
2. When is a Linear Program *bounded*?
   - Is there a minimum? Or is the minimum $-\infty$?

Important Questions

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

1. When is a Linear Program *feasible*?
   - Is there a solution to the constraints at all?
2. When is a Linear Program *bounded*?
   - Is there a minimum? Or is the minimum $-\infty$?
3. Can we characterize *optimality*?
   - How can we know that we found a minimum solution?
   - Do these solutions have nice description?
   - Do the solutions have *small bit complexity*?

Important Questions

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

1. When is a Linear Program *feasible*?
   - Is there a solution to the constraints at all?
2. When is a Linear Program *bounded*?
   - Is there a minimum? Or is the minimum $-\infty$?
3. Can we characterize *optimality*?
   - How can we know that we found a minimum solution?
   - Do these solutions have nice description?
   - Do the solutions have *small bit complexity*?
4. How do we design *efficient algorithms* that find *optimal solutions* to Linear Programs? interior point methods & ellipsoid

- Part I
  - Why Linear Programming?
  - Structural Results on Linear Programming
  - Duality Theory
- Part II
  - Game Theory
  - Learning Theory - Boosting
- Conclusion
- Acknowledgements

Fundamental Theorem of Linear Inequalities

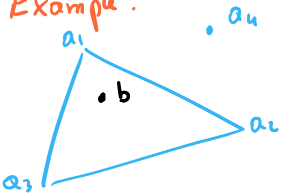
Theorem (Farkas (1894, 1898), Minkowski (1896))

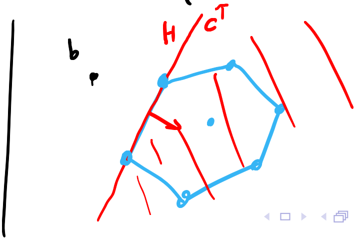
Let $a_1, \ldots, a_m, b \in \mathbb{R}^n$, and $t := \operatorname{rank}\{a_1, \ldots, a_m, b\}$. Then either

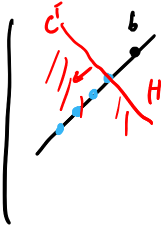
1. *b is a non-negative linear combination of linearly independent vectors from $a_1, \ldots, a_m$, or*
2. *there exists a hyperplane $H := \{x \mid c^T x = 0\}$ s.t.*
   - $c^T b < 0$
   - $c^T a_i \geq 0$
   - *H contains $t - 1$ linearly independent vectors from $a_1, \ldots, a_m$*

$\operatorname{rank}\{v_1, \ldots, v_x\}$ = max number of linearly independent vectors in $\{v_1, \ldots, v_x\}$

Example:







Fundamental Theorem of Linear Inequalities

Theorem (Farkas (1894, 1898), Minkowski (1896))

Let $a_1, \ldots, a_m, b \in \mathbb{R}^n$, *and* $t := \text{rank}\{a_1, \ldots, a_m, b\}$. *Then either*

1. *b is a non-negative linear combination of linearly independent vectors from $a_1, \ldots, a_m$, or*
2. *there exists a hyperplane $H := \{x \mid c^T x = 0\}$ s.t.*
   - $c^T b < 0$
   - $c^T a_i \geq 0$
   - *H contains $t - 1$ linearly independent vectors from $a_1, \ldots, a_m$*

Remark

The hyperplane H above is known as the *separating hyperplane*.

Farkas' Lemma

Lemma (Farkas Lemma)

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The following are equivalent:

1. *There exists $x \in \mathbb{R}^n$ such that $x \geq 0$ and $Ax = b$* (our LP feasible)
2. *$y^T b \geq 0$ for each $y \in \mathbb{R}^m$ such that $y^T A \geq 0$*

① ⇒ ② $\quad y^T A \geq 0 \quad Ax = b,\ x \geq 0 \quad \underbrace{\underbrace{y^T A}_{\geq 0}\underbrace{x}_{\geq 0}}_{\geq 0} = y^T b \Rightarrow y^T b \geq 0.$

② ⇒ ① (not ① ⇒ not ②)

$A_1, \dots, A_n$ columns of A, $A_1, \dots, A_n, b \in \mathbb{R}^m$

$\nexists\, x \geq 0$ s.t. $\sum_{i=1}^{n} x_i A_i = b$

(i.e. b is not nonnegative combination of $A_1, \dots, A_n$)

Fundamental thm of linear inequalities ⇒ then must have a separating hyperplane H_y s.t. $y^T A_i \geq 0\ \forall i$ and $\boxed{y^T b < 0}$ ⇒ not ②

$y^T A_i \geq 0\ \forall i \iff \boxed{y^T A \geq 0}$

Farkas' Lemma

Lemma (Farkas Lemma)

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. The following are equivalent:

1. *There exists $x \in \mathbb{R}^n$ such that $x \geq 0$ and $Ax = b$*
2. *$y^T b \geq 0$ for each $y \in \mathbb{R}^m$ such that $y^T A \geq 0$*

Equivalent formulation

Lemma (Farkas Lemma)

Let $A \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$. Then exactly one of the following statements hold:

1. *There exists $x \in \mathbb{R}^n$ such that $x \geq 0$ and $Ax = b$*
2. *There exists $y \in \mathbb{R}^m$ such that $y^T b > 0$ and $y^T A \leq 0$*

Linear Programming Duality

Consider our linear program:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Linear Programming Duality

Consider our linear program:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- From Farkas' lemma, we saw that $\overbrace{Ax = b \text{ and } x \geq 0}^{\text{LP feasible}}$ has a solution iff $y^T b \geq 0$ for each $y \in \mathbb{R}^m$ such that $y^T A \geq 0$.

Linear Programming Duality

Consider our linear program:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- From Farkas' lemma, we saw that $Ax = b$ and $x \geq 0$ has a solution iff $y^T b \geq 0$ for each $y \in \mathbb{R}^m$ such that $y^T A \geq 0$.
- If we look at what happens when we multiply $y^T A$, note the following:

$$\begin{aligned} y^T A \leq c^T &\Rightarrow y^T Ax \leq c^T x \\ &\Rightarrow y^T b \leq c^T x \end{aligned}$$

Linear Programming Duality

Consider our linear program:

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

- From Farkas' lemma, we saw that $Ax = b$ and $x \geq 0$ has a solution iff $y^T b \geq 0$ for each $y \in \mathbb{R}^m$ such that $y^T A \geq 0$.
- If we look at what happens when we multiply $y^T A$, note the following:

$$\begin{aligned} y^T A \leq c^T &\Rightarrow y^T A x \leq c^T x \\ &\Rightarrow y^T b \leq c^T x \end{aligned}$$

- Thus, if $y^T A \leq c^T$, then we have that $y^T b$ is a *lower bound* on the solution to our linear program!

Linear Programming Duality

Consider the following linear programs:

Primal LP

$$\begin{array}{ll} \text{minimize} & c^T x \\ \text{subject to} & Ax = b \\ & x \geq 0 \end{array}$$

Dual LP

best

lower bound

$$\begin{array}{ll} \text{maximize} & y^T b \\ \text{subject to} & \underbrace{y^T A \leq c^T}_{\text{constraint}} \end{array}$$

dual LP lower bounds the objective function of the primal LP.

Linear Programming Duality

Consider the following linear programs:

Primal LP	*Dual LP*
minimize $c^T x$	maximize $y^T b$
subject to $Ax = b$	subject to $y^T A \leq c^T$
$x \geq 0$	

- From previous slide

$$y^T A \leq c^T \Rightarrow y^T b \text{ is a lower bound on value of Primal}$$

Linear Programming Duality

Consider the following linear programs:

Primal LP	*Dual LP*
minimize $c^T x$	maximize $y^T b$
subject to $Ax = b$	subject to $y^T A \leq c^T$
$x \geq 0$	

- From previous slide

$$y^T A \leq c^T \Rightarrow y^T b \text{ is a lower bound on value of Primal}$$

- Thus, the optimal (maximum) value of *dual LP* lower bounds the optimal (minimum) value of the *Primal LP*!

Linear Programming Duality

Consider the following linear programs:

Primal LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual LP

$$\begin{aligned} \text{maximize} \quad & y^T b \\ \text{subject to} \quad & y^T A \leq c^T \end{aligned}$$

- From previous slide

$$y^T A \leq c^T \Rightarrow y^T b \text{ is a lower bound on value of Primal}$$

- Thus, the optimal (maximum) value of *dual LP* lower bounds the optimal (minimum) value of the *Primal LP*!

Theorem (Weak Duality)

Let x be a feasible solution of the primal LP and y be a feasible solution of the dual LP. Then

$$y^T b \leq c^T x.$$

Remarks on Duality

Primal LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual LP

$$\begin{aligned} \text{maximize} \quad & y^T b \\ \text{subject to} \quad & y^T A \leq c^T \end{aligned}$$

Remarks on Duality

Primal LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual LP

$$\begin{aligned} \text{maximize} \quad & y^T b \\ \text{subject to} \quad & y^T A \leq c^T \end{aligned}$$

- Optimal (maximum) value of *dual LP* lower bounds the optimal (minimum) value of the *Primal LP*!

Remarks on Duality

Primal LP		*Dual LP*	
$\alpha^* :=$ minimize	$c^T x$	$\beta^* =$ maximize	$y^T b$
subject to	$Ax = b$	subject to	$y^T A \leq c^T$
	$x \geq 0$		

- Optimal (maximum) value of *dual LP* lower bounds the optimal (minimum) value of the *Primal LP*!
- If $\alpha^*, \beta^* \in \mathbb{R}$ are the optimal values for primal and dual, respectively.

Remarks on Duality

Primal LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual LP

$$\begin{aligned} \text{maximize} \quad & y^T b \\ \text{subject to} \quad & y^T A \leq c^T \end{aligned}$$

- Optimal (maximum) value of *dual LP* lower bounds the optimal (minimum) value of the *Primal LP*!
- If $\alpha^*, \beta^* \in \mathbb{R}$ are the optimal values for primal and dual, respectively.
    - We showed that when both primal and dual are feasible, we have

$$\text{max dual } = \beta^* \leq \alpha^* = \text{ min of primal}$$

Remarks on Duality

Primal LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual LP

$$\begin{aligned} \text{maximize} \quad & y^T b \\ \text{subject to} \quad & y^T A \leq c^T \end{aligned}$$

- Optimal (maximum) value of *dual LP* lower bounds the optimal (minimum) value of the *Primal LP*!
- If $\alpha^*, \beta^* \in \mathbb{R}$ are the optimal values for primal and dual, respectively.
  - We showed that when both primal and dual are feasible, we have

$$\text{max dual } = \beta^* \leq \alpha^* = \text{ min of primal}$$

  - if primal *unbounded* ($\alpha^* = -\infty$) then dual *infeasible* ($\beta^* = -\infty$)

Remarks on Duality

Primal LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual LP

$$\begin{aligned} \text{maximize} \quad & y^T b \\ \text{subject to} \quad & y^T A \leq c^T \end{aligned}$$

- Optimal (maximum) value of *dual LP* lower bounds the optimal (minimum) value of the *Primal LP*!
- If $\alpha^*, \beta^* \in \mathbb{R}$ are the optimal values for primal and dual, respectively.
  - We showed that when both primal and dual are feasible, we have

    $$\text{max dual } = \beta^* \leq \alpha^* = \text{ min of primal}$$

  - if primal *unbounded* ($\alpha^* = -\infty$) then dual *infeasible* ($\beta^* = -\infty$)
  - if dual *unbounded* ($\beta^* = \infty$) then primal *infeasible* ($\alpha^* = \infty$)

Remarks on Duality

Primal LP		*Dual LP*	
minimize	$c^T x$	maximize	$y^T b$
subject to	$Ax = b$	subject to	$y^T A \leq c^T$
	$x \geq 0$		

- Optimal (maximum) value of *dual LP* lower bounds the optimal (minimum) value of the *Primal LP*!
- If $\alpha^*, \beta^* \in \mathbb{R}$ are the optimal values for primal and dual, respectively.
  - We showed that when both primal and dual are feasible, we have

$$\text{max dual } = \beta^* \leq \alpha^* = \text{ min of primal}$$

  - if primal *unbounded* ($\alpha^* = -\infty$) then dual *infeasible* ($\beta^* = -\infty$)
  - if dual *unbounded* ($\beta^* = \infty$) then primal *infeasible* ($\alpha^* = \infty$)
- **Practice problem:** show that dual of the dual LP is the primal LP!

Remarks on Duality

Primal LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual LP

$$\begin{aligned} \text{maximize} \quad & y^T b \\ \text{subject to} \quad & y^T A \leq c^T \end{aligned}$$

- Optimal (maximum) value of *dual LP* lower bounds the optimal (minimum) value of the *Primal LP*!
- If $\alpha^*, \beta^* \in \mathbb{R}$ are the optimal values for primal and dual, respectively.
  - We showed that when both primal and dual are feasible, we have

$$\text{max dual } = \beta^* \leq \alpha^* = \text{ min of primal}$$

  - if primal *unbounded* ($\alpha^* = -\infty$) then dual *infeasible* ($\beta^* = -\infty$)
  - if dual *unbounded* ($\beta^* = \infty$) then primal *infeasible* ($\alpha^* = \infty$)
- **Practice problem:** show that dual of the dual LP is the primal LP!
- When is the above inequality tight?

Strong Duality

Primal LP

$$\begin{aligned} \text{minimize} \quad & c^T x \\ \text{subject to} \quad & Ax = b \\ & x \geq 0 \end{aligned}$$

Dual LP

$$\begin{aligned} \text{maximize} \quad & y^T b \\ \text{subject to} \quad & y^T A \leq c^T \end{aligned}$$

- let $\alpha^*, \beta^* \in \mathbb{R}$ be optimal values for primal and dual, respectively.

Strong Duality

Primal LP	
minimize	$c^T x$
subject to	$Ax = b$
	$x \geq 0$

Dual LP	
maximize	$y^T b$
subject to	$y^T A \leq c^T$

- let $\alpha^*, \beta^* \in \mathbb{R}$ be optimal values for primal and dual, respectively.

Theorem (Strong Duality)

If primal LP or dual LP is feasible, then

$$\text{max dual} = \boxed{\beta^* = \alpha^*} = \text{min of primal.}$$

always equality

Proof of Strong Duality

Theorem (Strong Duality)

If primal LP or dual LP is feasible, then

$$\textit{max dual} = \beta^* = \alpha^* = \textit{min of primal}.$$

Proof of Strong Duality

Theorem (Strong Duality)

If primal LP or dual LP is feasible, then

$$\textit{max dual } = \beta^* = \alpha^* = \textit{ min of primal.}$$

1. Let x^* be such that $c^T x^* = \alpha^*$. Can assume that $\alpha^* \neq -\infty$.

Proof of Strong Duality

Theorem (Strong Duality)

If primal LP or dual LP is feasible, then

$$\text{max dual } = \beta^* = \alpha^* = \text{ min of primal.}$$

1. Let x^* be such that $c^T x^* = \alpha^*$. Can assume that $\alpha^* \neq -\infty$.
2. Let $B = \underbrace{\begin{pmatrix} A \\ -c^T \end{pmatrix}}_{\mathbb{R}^{(m+1)\times n}}$ and $v(\varepsilon) = \begin{pmatrix} b \\ \underbrace{-\alpha^* + \varepsilon}_{-(\alpha^* - \varepsilon)} \end{pmatrix}$ $\quad \varepsilon \geq 0$

$\alpha^* - \varepsilon \leq \alpha^* = \text{min primal}$

Proof of Strong Duality

Theorem (Strong Duality)

If primal LP or dual LP is feasible, then

$$\text{max dual} = \beta^* = \alpha^* = \text{min of primal}.$$

1. Let x^* be such that $c^T x^* = \alpha^*$. Can assume that $\alpha^* \neq -\infty$.
2. Let $B = \begin{pmatrix} A \\ -c^T \end{pmatrix}$ and $v(\varepsilon) = \begin{pmatrix} b \\ -\alpha^* + \varepsilon \end{pmatrix}$
3. Apply Farkas' lemma on $Bx = v(0)$ and $x \geq 0$. This system has a solution, so we get:

x^*

$$(y^T \; z) B \leq 0 \Rightarrow (y^T \; z)\begin{pmatrix} b \\ -\alpha^* \end{pmatrix} \leq 0 \Leftrightarrow y^T b - z\alpha^* \leq 0$$

$$y^T A - zc^T \leq 0 \Rightarrow y^T b - z\alpha^* \leq 0$$

$$\boxed{y^T A \leq zc^T \Rightarrow y^T b \leq z\alpha^*}$$

if $z = 0$ then we have $y^T A \leq 0 \Rightarrow y^T b \leq 0$.

Proof of Strong Duality

Theorem (Strong Duality)

If primal LP or dual LP is feasible, then

$$\text{max dual} = \beta^* = \alpha^* = \text{min of primal.}$$

1. Let x^* be such that $c^T x^* = \alpha^*$. Can assume that $\alpha^* \neq -\infty$.
2. Let $B = \begin{pmatrix} A \\ -c^T \end{pmatrix}$ and $v(\varepsilon) = \begin{pmatrix} b \\ -\alpha^* + \varepsilon \end{pmatrix}$
3. Apply Farkas' lemma on $Bx = v(0)$ and $x \geq 0$. This system has a solution, so we get:
4. Now, if $\varepsilon > 0$, applying Farkas' lemma on system $Bx = v(\varepsilon)$ and $x \geq 0$ we get: $\underset{x \geq 0}{Bx = v(\varepsilon)}$ has no solution (by optimality of α^*)

Farkas lemma: there is $(y^T \; z) \in \mathbb{R}^{m+1}$ s.t. $y^T A \leq c^T z$ and $y^T b > z(\alpha^* - \varepsilon)$

by previous slide $z \neq 0$ (can assume that $z > 0$)

$\Rightarrow \exists\, y_\varepsilon$ s.t. $y_\varepsilon^T b > \alpha^* - \varepsilon$ $(\forall \varepsilon)$

Proof of Strong Duality

Theorem (Strong Duality)

If primal LP or dual LP is feasible, then

$$\text{max dual} \; = \beta^* = \alpha^* = \; \text{min of primal.}$$

1. Let x^* be such that $c^T x^* = \alpha^*$. Can assume that $\alpha^* \neq -\infty$.
2. Let $B = \begin{pmatrix} A \\ -c^T \end{pmatrix}$ and $v(\varepsilon) = \begin{pmatrix} b \\ -\alpha^* + \varepsilon \end{pmatrix}$
3. Apply Farkas' lemma on $Bx = v(0)$ and $x \geq 0$. This system has a solution, so we get:
4. Now, if $\varepsilon > 0$, applying Farkas' lemma on system $Bx = v(\varepsilon)$ and $x \geq 0$ we get:
5. Thus, for any $\varepsilon > 0$ there is $y \in \mathbb{R}^m$ such that $y^T A \leq c^T$ and $\beta^* \geq y^T b > \alpha^* - \varepsilon$.

y feasible solution

$\Rightarrow \boxed{\beta^* = \alpha^*}$

Affine form of Farkas' Lemma

A consequence of LP duality is the following lemma:

Lemma (Affine Farkas' Lemma)

Let the system

$$Ax \leq b$$

have at least one solution, and suppose that inequality

$$c^T x \leq \delta$$

holds whenever x *satisfies* $Ax \leq b$. *Then, for some* $\delta' \leq \delta$ *the linear inequality*

$$c^T x \leq \delta'$$

is a non-negative linear combination of the inequalities of $Ax \leq b$.

Affine form of Farkas' Lemma

A consequence of LP duality is the following lemma:

Lemma (Affine Farkas' Lemma)

Let the system

$$Ax \leq b$$

have at least one solution, and suppose that inequality

$$c^T x \leq \delta$$

holds whenever x satisfies $Ax \leq b$. Then, for some $\delta' \leq \delta$ the linear inequality

$$c^T x \leq \delta'$$

is a non-negative linear combination of the inequalities of $Ax \leq b$.

Practice problem: use LP duality and Farkas' lemma to prove this lemma!

Two-player games

Setup:

- Two players (Alice and Bob)
- Each player has a (finite) set of strategies $S_A = \{1, \ldots, m\}$ and $S_B = \{1, \ldots, n\}$

Two-player games

Setup:

- Two players (Alice and Bob)
- Each player has a (finite) set of strategies $S_A = \{1, \ldots, m\}$ and $S_B = \{1, \ldots, n\}$
- Payoff matrices $A, B \in \mathbb{R}^{m \times n}$ for Alice and Bob, respectively

Two-player games

Setup:

- Two players (Alice and Bob)
- Each player has a (finite) set of strategies $S_A = \{1, \ldots, m\}$ and $S_B = \{1, \ldots, n\}$
- Payoff matrices $A, B \in \mathbb{R}^{m \times n}$ for Alice and Bob, respectively
  - If Alice plays i and Bob plays j, then
  - Alice gets A_{ij}
  - Bob gets B_{ij}

Two-player games

Setup:

- Two players (Alice and Bob)
- Each player has a (finite) set of strategies $S_A = \{1, \ldots, m\}$ and $S_B = \{1, \ldots, n\}$
- Payoff matrices $A, B \in \mathbb{R}^{m \times n}$ for Alice and Bob, respectively
    - If Alice plays i and Bob plays j, then
    - Alice gets A_{ij}
    - Bob gets B_{ij}
- Example: *battle of the sexes* game

Two-player games

Setup:

- Two players (Alice and Bob)
- Each player has a (finite) set of strategies $S_A = \{1, \ldots, m\}$ and $S_B = \{1, \ldots, n\}$
- Payoff matrices $A, B \in \mathbb{R}^{m \times n}$ for Alice and Bob, respectively
  - If Alice plays i and Bob plays j, then
  - Alice gets A_{ij}
  - Bob gets B_{ij}
- Example: *battle of the sexes* game

Bob

$$\begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

Alice

$$\begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}$$

Bob

Alice

	Football	Opera
Football	$\boxed{(2,1)}$	(0,0)
Opera	(0,0)	$\boxed{(1,2)}$

Table: Battle of the sexes payoff matrices

Nash Equilibrium

Assuming players are rational, i.e. want to maximize their payoffs, we have:

Definition (Nash Equilibrium)

A strategy profile (i, j) is called a Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $A_{ij} \geq A_{kj}$ for all $k \in S_A$
2. $B_{ij} \geq B_{i\ell}$ for all $\ell \in S_B$

1. if Alice knew Bob playing j
then Alice has no incentive to not play i

Nash Equilibrium

Assuming players are rational, i.e. want to maximize their payoffs, we have:

Definition (Nash Equilibrium)

A strategy profile (i, j) is called a Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $A_{ij} \geq A_{kj}$ for all $k \in S_A$
2. $B_{ij} \geq B_{i\ell}$ for all $\ell \in S_B$

	Football	Opera
Football	(2,1)	(0,0)
Opera	(0,0)	(1,2)

Table: Battle of the sexes payoff matrices

Nash Equilibrium

Assuming players are rational, i.e. want to maximize their payoffs, we have:

Definition (Nash Equilibrium)

A strategy profile (i, j) is called a Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $A_{ij} \geq A_{kj}$ for all $k \in S_A$
2. $B_{ij} \geq B_{i\ell}$ for all $\ell \in S_B$

	Football	Opera
Football	(2,1)	(0,0)
Opera	(0,0)	(1,2)

Table: Battle of the sexes payoff matrices

Bob

Alice

	Silent	Snitch
Silent	(-1,-1)	(-10,0)
Snitch	(0,-10)	$\boxed{(-5,-5)}$

Nash equilibrium

Table: Prisoner's dilemma

Mixed Strategies

Definition (Mixed Strategy)

A mixed strategy is a probability distribution over a set of pure strategies S. If Alice's strategies are $S_A = \{1, \ldots, n\}$, her mixed strategies are:

$$\Delta_A := \{x \in \mathbb{R}^n \mid \underbrace{x \geq 0 \text{ and } \|x\|_1 = 1}_{x \text{ is a probability distribution}}\}$$

$x_i \leftarrow \Pr[\text{Alice plays } i]$

Mixed Strategies

Definition (Mixed Strategy)

A mixed strategy is a probability distribution over a set of pure strategies S. If Alice's strategies are $S_A = \{1, \dots, n\}$, her mixed strategies are:

$$\Delta_A := \{x \in \mathbb{R}^n \mid x \geq 0 \text{ and } \|x\|_1 = 1\}$$

- Models situation where players choose their strategy "at random"

Mixed Strategies

Definition (Mixed Strategy)

A mixed strategy is a probability distribution over a set of pure strategies S. If Alice's strategies are $S_A = \{1, \ldots, n\}$, her mixed strategies are:

$$\Delta_A := \{x \in \mathbb{R}^n \mid x \geq 0 \text{ and } \|x\|_1 = 1\}$$

- Models situation where players choose their strategy "at random"
- Payoffs for each player defined as *expected gain*. That is, (x, y) is the profile of mixed strategies used by Alice and Bob, we have:

Mixed Strategies

Definition (Mixed Strategy)

A mixed strategy is a probability distribution over a set of pure strategies S. If Alice's strategies are $S_A = \{1, \ldots, n\}$, her mixed strategies are:

$$\Delta_A := \{x \in \mathbb{R}^n \mid x \geq 0 \text{ and } \|x\|_1 = 1\}$$

- Models situation where players choose their strategy "at random"
- Payoffs for each player defined as *expected gain*. That is, (x, y) is the profile of mixed strategies used by Alice and Bob, we have:

$$v_A(x, y) = \sum_{(i,j) \in S_A \times S_B} A_{ij} \underbrace{x_i y_j} = x^T A y$$

$$v_B(x, y) = \sum_{(i,j) \in S_A \times S_B} B_{ij} \underbrace{x_i y_j} = x^T B y$$

Nash Equilibrium Mixed Strategies

Assuming players are rational, i.e. want to maximize their payoffs, we have:

Definition ((Mixed) Nash Equilibrium)

A strategy profile $x \in \Delta_A, y \in \Delta_B$ is called a (mixed) Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $x^T Ay \geq z^T Ay$ for all $z \in \Delta_A$
2. $x^T By \geq x^T Bw$ for all $w \in \Delta_B$

Nash Equilibrium Mixed Strategies

Assuming players are rational, i.e. want to maximize their payoffs, we have:

Definition ((Mixed) Nash Equilibrium)

A strategy profile $x \in \Delta_A, y \in \Delta_B$ is called a (mixed) Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $x^T Ay \geq z^T Ay$ for all $z \in \Delta_A$
2. $x^T By \geq x^T Bw$ for all $w \in \Delta_B$

Goalie

'Player

	Jump left	Jump right
kick left	(-1,1)	(1,-1)
kick right	(1,-1)	(-1,1)

Table: Penalty Kick

no pure NE (player always has incentive to deviate)

Nash Equilibrium Mixed Strategies

Assuming players are rational, i.e. want to maximize their payoffs, we have:

Definition ((Mixed) Nash Equilibrium)

A strategy profile $x \in \Delta_A, y \in \Delta_B$ is called a (mixed) Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $x^T Ay \geq z^T Ay$ for all $z \in \Delta_A$
2. $x^T By \geq x^T Bw$ for all $w \in \Delta_B$

	Jump left	Jump right
kick left	(-1,1)	(1,-1)
kick right	(1,-1)	(-1,1)

Table: Penalty Kick

- *Zero-Sum Game*: payoff matrices satisfy $A = -B$

$\boxed{A+B=0}$

Nash Equilibrium Mixed Strategies

Assuming players are rational, i.e. want to maximize their payoffs, we have:

Definition ((Mixed) Nash Equilibrium)

A strategy profile $x \in \Delta_A, y \in \Delta_B$ is called a (mixed) Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $x^T Ay \geq z^T Ay$ for all $z \in \Delta_A$
2. $x^T By \geq x^T Bw$ for all $w \in \Delta_B$

	Jump left	Jump right
kick left	(-1,1)	(1,-1)
kick right	(1,-1)	(-1,1)

Table: Penalty Kick

- *Zero-Sum Game*: payoff matrices satisfy $A = -B$
- No pure Nash Equilibrium! Practice problem!

Nash Equilibrium Mixed Strategies

Assuming players are rational, i.e. want to maximize their payoffs, we have:

Definition ((Mixed) Nash Equilibrium)

A strategy profile $x \in \Delta_A, y \in \Delta_B$ is called a (mixed) Nash equilibrium if the strategy played by each player is optimal, *given the strategy of the other player*. That is:

1. $x^T Ay \geq z^T Ay$ for all $z \in \Delta_A$
2. $x^T By \geq x^T Bw$ for all $w \in \Delta_B$

	Jump left	Jump right
kick left	(-1,1)	(1,-1)
kick right	(1,-1)	(-1,1)

Table: Penalty Kick

- *Zero-Sum Game*: payoff matrices satisfy $A = -B$
- No pure Nash Equilibrium!
- One mixed Nash equilibrium: $x = y = (1/2, 1/2)$

Practice problem!

Von Neumann's Minimax Theorem

Theorem

In a zero-sum game, for any payoff matrix $A \in \mathbb{R}^{m \times n}$:

$$\alpha \quad \underbrace{\max_{x \in \Delta_A}}_{\text{Alice wants to max her payoff}} \min_{y \in \Delta_B} \underbrace{x^T A y}_{\text{Alice's payoff}} = \underbrace{\min_{y \in \Delta_B} \max_{x \in \Delta_A} x^T A y}_{\text{Alice's payoff once Bob tells his strategy}} \quad \beta$$

Choose x so that whatever Bob plays Alice gets still α

LHS: Alice picks first her strategy

RHS: Bob picks first his strategy

Von Neumann's Minimax Theorem

Theorem

In a zero-sum game, for any payoff matrix $A \in \mathbb{R}^{m \times n}$:

$$\max_{x \in \Delta_A} \min_{y \in \Delta_B} x^T A y = \min_{y \in \Delta_B} \max_{x \in \Delta_A} x^T A y$$

For given $x \in \Delta_A$:

$$\min_{y \in \Delta_B} x^T A y = \min_{j \in S_B} (x^T A)_j$$

Von Neumann's Minimax Theorem

Theorem

In a zero-sum game, for any payoff matrix $A \in \mathbb{R}^{m \times n}$:

$$\max_{x \in \Delta_A} \min_{y \in \Delta_B} x^T A y = \min_{y \in \Delta_B} \max_{x \in \Delta_A} x^T A y$$

For given $x \in \Delta_A$:

$$\min_{y \in \Delta_B} x^T A y = \min_{j \in S_B} (x^T A)_j$$

Left hand side can be written as

$$\begin{aligned}
\max \quad & s \\
\text{s.t.} \quad & s \leq (x^T A)_j \quad \text{for } j \in S_B \\
& \sum_{i \in S_A} x_i = 1 \\
& x \geq 0
\end{aligned}$$

Von Neumann's Minimax Theorem

Theorem

In a zero-sum game, for any payoff matrix $A \in \mathbb{R}^{m\times n}$:

$$\max_{x\in\Delta_A} \min_{y\in\Delta_B} x^T Ay = \min_{y\in\Delta_B} \max_{x\in\Delta_A} x^T Ay$$

For given $x \in \Delta_A$:

$$\min_{y\in\Delta_B} x^T Ay = \min_{j\in S_B} (x^T A)_j$$

For given $y \in \Delta_B$:

$$\max_{x\in\Delta_A} x^T Ay = \max_{i\in S_A} (Ay)_i$$

Left hand side can be written as

$$\begin{aligned}
\max \quad & s \\
\text{s.t.} \quad & s \leq (x^T A)_j \quad \text{for } j \in S_B \\
& \sum_{i\in S_A} x_i = 1 \\
& x \geq 0
\end{aligned}$$

Von Neumann's Minimax Theorem

Theorem

In a zero-sum game, for any payoff matrix $A \in \mathbb{R}^{m \times n}$:

$$\max_{x \in \Delta_A} \min_{y \in \Delta_B} x^T A y = \min_{y \in \Delta_B} \max_{x \in \Delta_A} x^T A y$$

For given $x \in \Delta_A$:

pure strategy

$$\min_{y \in \Delta_B} x^T A y = \min_{j \in S_B} (x^T A)_j$$

Left hand side can be written as

$$\begin{aligned} \max \quad & s \\ \text{s.t.} \quad & s \leq (x^T A)_j \quad \text{for } j \in S_B \\ & \sum_{i \in S_A} x_i = 1 \\ & x \geq 0 \end{aligned}$$

any pure strategy for Bob

over all Alice's strategy

For given $y \in \Delta_B$:

$$\max_{x \in \Delta_A} x^T A y = \max_{i \in S_A} (A y)_i$$

Right hand side can be written as

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & t \geq (A y)_i \quad \text{for } i \in S_A \\ & \sum_{j \in S_B} y_j = 1 \\ & y \geq 0 \end{aligned}$$

Von Neumann's Minimax Theorem

Theorem

In a zero-sum game, for any payoff matrix $A \in \mathbb{R}^{m \times n}$:

$$\max_{x \in \Delta_A} \min_{y \in \Delta_B} x^T A y = \min_{y \in \Delta_B} \max_{x \in \Delta_A} x^T A y$$

These two LPs are a primal-dual pair! ($\Rightarrow$ strong duality)

Left hand side can be written as

$$\begin{aligned} \max \quad & s \\ \text{s.t.} \quad & s \leq (x^T A)_j \quad \text{for } j \in S_B \\ & \sum_{i \in S_A} x_i = 1 \\ & x \geq 0 \end{aligned}$$

Right hand side can be written as

$$\begin{aligned} \min \quad & t \\ \text{s.t.} \quad & t \geq (Ay)_i \quad \text{for } i \in S_A \\ & \sum_{j \in S_B} y_j = 1 \\ & y \geq 0 \end{aligned}$$

- Part I
  - Why Linear Programming?
  - Structural Results on Linear Programming
  - Duality Theory
- Part II
  - Game Theory
  - Learning Theory - Boosting
- Conclusion
- Acknowledgements

Learning Theory

Consider classification problem over $\mathcal{X}$:

Learning Theory

Consider classification problem over $\mathcal{X}$:

- Set of hypothesis $\mathcal{H} := \{h : \mathcal{X} \to \{0, 1\}\}$

Learning Theory

Consider classification problem over $\mathcal{X}$:

- Set of hypothesis $\mathcal{H} := \{h : \mathcal{X} \to \{0, 1\}\}$
- Each $x \in \mathcal{X}$ has a correct value $c(x) \in \{0, 1\}$

Learning Theory

Consider classification problem over $\mathcal{X}$:

- Set of hypothesis $\mathcal{H} := \{h : \mathcal{X} \to \{0, 1\}\}$
- Each $x \in \mathcal{X}$ has a correct value $c(x) \in \{0, 1\}$
- Data is sampled from unknown distribution $q \in \underbrace{\Delta_{\mathcal{X}}}$

unknown
distribution
over elements

Learning Theory

Consider classification problem over $\mathcal{X}$:

- Set of hypothesis $\mathcal{H} := \{h : \mathcal{X} \to \{0,1\}\}$
- Each $x \in \mathcal{X}$ has a correct value $c(x) \in \{0,1\}$
- Data is sampled from unknown distribution $q \in \Delta_{\mathcal{X}}$
- *Weak learning assumption*:

For any distribution $q \in \Delta_{\mathcal{X}}$, there is a hypothesis $h \in \mathcal{H}$ which is wrong less than half the time.

$$\underbrace{\exists \gamma > 0}_{\text{gap}},\ \underbrace{\forall q \in \Delta_{\mathcal{X}}}_{\text{any dist.}},\ \underbrace{\exists h \in \mathcal{H}}_{\text{there is hypothesis}},\quad \underbrace{\Pr_{x\sim q}[h(x) \neq c(x)]}_{\text{Pr } h \text{ being wrong}} \leq \underbrace{\frac{1-\gamma}{2}}_{\text{less than } 1/2 \text{ the time (by gap } \gamma)}$$

Learning Theory

Consider classification problem over $\mathcal{X}$:

- Set of hypothesis $\mathcal{H} := \{h : \mathcal{X} \to \{0,1\}\}$
- Each $x \in \mathcal{X}$ has a correct value $c(x) \in \{0,1\}$
- Data is sampled from unknown distribution $q \in \Delta_{\mathcal{X}}$
- *Weak learning assumption*:

  For any distribution $q \in \Delta_{\mathcal{X}}$, there is a hypothesis $h \in \mathcal{H}$ which is wrong less than half the time.

  $$\exists \gamma > 0,\ \forall q \in \Delta_{\mathcal{X}},\ \exists h \in \mathcal{H}, \quad \Pr_{x \sim q}[h(x) \neq c(x)] \leq \frac{1-\gamma}{2}$$

- Surprisingly, weak learning assumption implies something much stronger: it is possible to *combine* classifiers in $\mathcal{H}$ to construct a *classifier* that is *always right* (known as *strong learning*).

Boosting

on set of hypotheses

Theorem

Let $\mathcal{H}$ be a set of hypotheses satisfying *weak learning assumption*. *Then there is distribution $p \in \Delta_{\mathcal{H}}$ such that the* *weighed majority classifier*

$$c_p(x) := \begin{cases} 1, & \text{if } \sum_{h \in \mathcal{H}} p_h \cdot h(x) \geq 1/2 \\ 0, & \text{otherwise} \end{cases}$$

is always correct. That is, $c_p(x) = c(x)$ for all $x \in \mathcal{X}$

Boosting

Theorem

Let $\mathcal{H}$ be a set of hypotheses satisfying weak learning assumption*. Then there is distribution $p \in \Delta_{\mathcal{H}}$ such that the* weighed majority classifier

$$c_p(x) := \begin{cases} 1, & \text{if } \sum_{h \in \mathcal{H}} p_h \cdot h(x) \geq 1/2 \\ 0, & \text{otherwise} \end{cases}$$

is always correct. That is, $c_p(x) = c(x)$ for all $x \in \mathcal{X}$

- Let $M \in \{-1, 1\}^{m \times n}$, where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if classifier } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

payoff matrix

classifiers

examples

Boosting

Theorem

Let $\mathcal{H}$ be a set of hypotheses satisfying weak learning assumption. Then there is distribution $p \in \Delta_{\mathcal{H}}$ such that the weighed majority classifier

$$c_p(x) := \begin{cases} 1, & \text{if } \sum_{h \in \mathcal{H}} p_h \cdot h(x) \geq 1/2 \\ 0, & \text{otherwise} \end{cases}$$

is always correct. That is, $c_p(x) = c(x)$ for all $x \in \mathcal{X}$

- Let $M \in \{-1, 1\}^{m \times n}$, where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if classifier } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

- Weak learning:

$q \in \Delta_X$

$$\sum_{1 \leq i \leq m} q_i \cdot \underbrace{\delta_{h_j(x_i) \neq c(x_i)}}_{} \leq \frac{1-\gamma}{2}$$

LHS: prob. that h is wrong

$\begin{cases} 1 & \text{if } h \text{ wrong on } x_i \\ 0 & \text{o.w.} \end{cases}$

Boosting - Proof

Let $M \in \{-1, 1\}^{m \times n}$,
where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

Weak learning:

$$\sum_{1 \le i \le n} q_j \cdot \delta_{h_j(x_i) \neq c(x_i)} \le \frac{1-\gamma}{2}$$

Boosting - Proof

Let $M \in \{-1, 1\}^{m \times n}$,
where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

Weak learning:

$$\sum_{1 \leq i \leq m} q_j \cdot \delta_{h_j(x_i) \neq c(x_i)} \leq \frac{1-\gamma}{2}$$

- Note that $M_{ij} = 2 \cdot \delta_{h_j(x_i) \neq c(x_i)} - 1$

$$\boxed{q^T M e_j \leq -\gamma} \quad \Rightarrow \quad q^T M p \leq -\gamma$$

for any $p \in \Delta_{\mathcal{H}}$.

$$\sum q_i \; 2 \cdot \delta_{i,j} \; - \underbrace{\sum q_i}_{1} \; \leq -\gamma$$

$$q^T M e_j = \sum q_i (2\delta_{ij} - 1) \leq -\gamma$$

Boosting - Proof

Let $M \in \{-1, 1\}^{m \times n}$,
where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

Weak learning:

$$\sum_{1 \le i \le n} q_j \cdot \delta_{h_j(x_i) \neq c(x_i)} \le \frac{1-\gamma}{2}$$

- Note that $M_{ij} = 2 \cdot \delta_{h_j(x_i) \neq c(x_i)} - 1$

$$q^T M e_j \le -\gamma \quad \Rightarrow \quad \underbrace{q^T M p}_{\text{value of our game}} \le -\gamma$$

  for any $p \in \Delta_{\mathcal{H}}$.
- By minimax, we have:

$$\max_{q \in \Delta_{\mathcal{X}}} \min_{p \in \Delta_{\mathcal{H}}} q^T M p = \underbrace{\min_{p \in \Delta_{\mathcal{H}}}}_{\text{classifiers}} \max_{q \in \Delta_{\mathcal{X}}} q^T M p \boxed{\le -\gamma}$$

Boosting - Proof

$\sum p_j \cdot h_j(x) \geq \frac{1+\gamma}{2} \Rightarrow c_p(x)=1$

Let $M \in \{-1,1\}^{m \times n}$,
where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

Weak learning:

$$\sum_{1 \leq i \leq n} q_j \cdot \delta_{h_j(x_i) \neq c(x_i)} \leq \frac{1-\gamma}{2}$$

$c(x)=1$

$\sum p_j \, \delta_{h_j(x)=0} \leq \frac{1-\gamma}{2}$

- Note that $M_{ij} = 2 \cdot \delta_{h_j(x_i) \neq c(x_i)} - 1$

$$q^T M e_j \leq -\gamma \quad \Rightarrow \quad q^T M p \leq -\gamma$$

for any $p \in \Delta_{\mathcal{H}}$.

$\sum p_j \, \delta_{h_j \text{ wrong}} \leq \frac{1-\gamma}{2}$

- By minimax, we have:

$$\max_{q \in \Delta_{\mathcal{X}}} \min_{p \in \Delta_{\mathcal{H}}} q^T M p = \boxed{\min_{p \in \Delta_{\mathcal{H}}} \max_{q \in \Delta_{\mathcal{X}}} q^T M p \leq -\gamma}$$

- In particular, right hand side implies weighted classifier *always* correct.

Boosting - Proof

Let $M \in \{-1, 1\}^{m \times n}$,
where $m = |\mathcal{X}|$ and $n = |\mathcal{H}|$.

$$M_{ij} = \begin{cases} +1, & \text{if } h_j \text{ wrong on } x_i \\ -1, & \text{otherwise} \end{cases}$$

Weak learning:

$$\sum_{1 \le i \le n} q_j \cdot \delta_{h_j(x_i) \neq c(x_i)} \le \frac{1-\gamma}{2}$$

- By minimax, we have:

$$\max_{q \in \Delta_{\mathcal{X}}} \min_{p \in \Delta_{\mathcal{H}}} q^T M p = \min_{p \in \Delta_{\mathcal{H}}} \max_{q \in \Delta_{\mathcal{X}}} q^T M p \le -\gamma$$

- In particular, right hand side implies weighted classifier *always* correct.

Conclusion

- Mathematical programming - very general, and pervasive in Algorithmic life

Conclusion

- Mathematical programming - very general, and pervasive in Algorithmic life
- General mathematical programming very hard (how hard do you think it is?)

Conclusion

- Mathematical programming - very general, and pervasive in Algorithmic life
- General mathematical programming very hard (how hard do you think it is?)
- Special cases have very striking applications!

Today: *Linear Programming*

Conclusion

- Mathematical programming - very general, and pervasive in Algorithmic life
- General mathematical programming very hard (how hard do you think it is?)
- Special cases have very striking applications!

Today: *Linear Programming*

- Linear Programming and Duality - fundamental concepts, lots of applications!

Conclusion

- Mathematical programming - very general, and pervasive in Algorithmic life
- General mathematical programming very hard (how hard do you think it is?)
- Special cases have very striking applications!

Today: *Linear Programming*

- Linear Programming and Duality - fundamental concepts, lots of applications!
  - Applications in Combinatorial Optimization (a lot of it happened here at UW!)
  - Applications in Game Theory (minimax theorem)
  - Applications in Learning Theory (boosting)
  - many more

Acknowledgement

- Lecture based largely on:
  - Lectures 3-6 of Yarom Singer's Advanced Optimization class
  - [Schrijver 1986, Chapter 7]
- See Yarom's notes at `https://people.seas.harvard.edu/~yaron/AM221-S16/schedule.html`

References I

Schirjver, Alexander (1986)
Theory of Linear and Integer Programming

Fourier, J. B. 1826
Analyse des travaux de l'Académie Royale des Sciences pendant l'année 1823.
Partie mathématique (1826)

Fourier, J. B. 1827
Analyse des travaux de l'Académie Royale des Sciences pendant l'année 1824.
Partie mathématique (1827)