Submodular Function Optimization

AU4606: Network Optimization

Xiaoming Duan
Department of Automation
Shanghai Jiao Tong University

December 11, 2023

Last few lectures

- Minimum spanning tree problems
    - Optimality conditions
    - Algorithms
        1. Kruskal's algorithm
        2. Prim's algorithm
        3. Sollin's algorithm
    - Matroids and greedy algorithms

Today

1. Submodular functions
2. Monotone submodular maximization
3. Unconstrained submodular minimization

Today

1. Submodular functions
2. Monotone submodular maximization
3. Unconstrained submodular minimization

Greedy works for matroid optimization

An optimization problem over matroids

Let $(E, \mathcal{I})$ be a matroid and $w : E \to \mathbb{R}$ be a function that assigns a cost to each element of E.
Define $w(X) = \sum_{x \in X} w(x)$ for any nonempty subset $X \subseteq E$.
Find a maximal independent set that has the minimum cost.

Greedy works

If $(E, \mathcal{I})$ is a matroid, then greedy algorithm finds a maximal independent set that has the minimum cost.

For problems with special structure, they may be solved efficiently.

Greedy can fail

Problem setup

- Let $\{1, 2, \ldots, 2k+1\}$ be a ground set
- The function value
  1. $f(\{e\}) = 1$ for $e \in \{1, \ldots, k\}$
  2. $f(\{e\}) = \epsilon$ for $e \in \{k+1, \ldots, 2k\}$
  3. $f(\{2k+1\}) = 1 + \epsilon$
  4. For $S_1 \subseteq \{1, \ldots, k\}$ and $S_2 \subseteq \{k+1, \ldots, 2k\}$, we have $f(S_1 \cup S_2) = |S_1| + \epsilon |S_2|$
  5. For $S_1 \subseteq \{1, \ldots, k\}$ and $S_2 \subseteq \{k+1, \ldots, 2k\}$, we have $f(S_1 \cup S_2 \cup \{2k+1\}) = \max\{|S_1|, 1 + \epsilon\} + \epsilon |S_2|$

Pick a k-element set S to maximize $f(S)$?

Greedy can fail

Problem setup

- Let $\{1, 2, \ldots, 2k+1\}$ be a ground set
- The function value
  1. $f(\{e\}) = 1$ for $e \in \{1, \ldots, k\}$
  2. $f(\{e\}) = \epsilon$ for $e \in \{k+1, \ldots, 2k\}$
  3. $f(\{2k+1\}) = 1 + \epsilon$
  4. For $S_1 \subseteq \{1, \ldots, k\}$ and $S_2 \subseteq \{k+1, \ldots, 2k\}$, we have $f(S_1 \cup S_2) = |S_1| + \epsilon |S_2|$
  5. For $S_1 \subseteq \{1, \ldots, k\}$ and $S_2 \subseteq \{k+1, \ldots, 2k\}$, we have $f(S_1 \cup S_2 \cup \{2k+1\}) = \max\{|S_1|, 1+\epsilon\} + \epsilon |S_2|$

Pick a k-element set S to maximize $f(S)$? Pick $\{1, \ldots, k\}$ and obtain k

Greedy can fail

Problem setup

- Let $\{1, 2, \ldots, 2k+1\}$ be a ground set
- The function value
  1. $f(\{e\}) = 1$ for $e \in \{1, \ldots, k\}$
  2. $f(\{e\}) = \epsilon$ for $e \in \{k+1, \ldots, 2k\}$
  3. $f(\{2k+1\}) = 1 + \epsilon$
  4. For $S_1 \subseteq \{1, \ldots, k\}$ and $S_2 \subseteq \{k+1, \ldots, 2k\}$, we have $f(S_1 \cup S_2) = |S_1| + \epsilon |S_2|$
  5. For $S_1 \subseteq \{1, \ldots, k\}$ and $S_2 \subseteq \{k+1, \ldots, 2k\}$, we have $f(S_1 \cup S_2 \cup \{2k+1\}) = \max\{|S_1|, 1+\epsilon\} + \epsilon |S_2|$

Pick a k-element set S to maximize $f(S)$? Pick $\{1, \ldots, k\}$ and obtain k

Greedy works as follows

- Pick $\{2k+1\}$ in the first iteration
- Pick $k-1$ elements from $\{k+1, \ldots, 2k\}$
- Obtain a function value of $\epsilon(k-1) + 1 + \epsilon = 1 + \epsilon k$

Set function optimization

We are often interested in problems of the following form

$$\max\{f(S) : S \in \mathcal{F}\}$$
$$\min\{f(S) : S \in \mathcal{F}\}$$

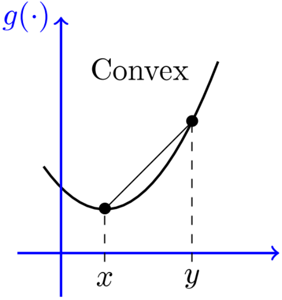
where $\mathcal{F} \subseteq 2^N$ is a collection of feasible subsets

Examples

1. min/max cut: $f(S)$ is cut value of cut set $(S, \bar{S})$ (N is node set)
2. min/max spanning tree: $f(S)$ is total arc weights ($\mathcal{F}$: spanning trees)
3. min vertex cover: $f(S) = |S|$, $\mathcal{F}$ is all sets of nodes that "cover" arcs

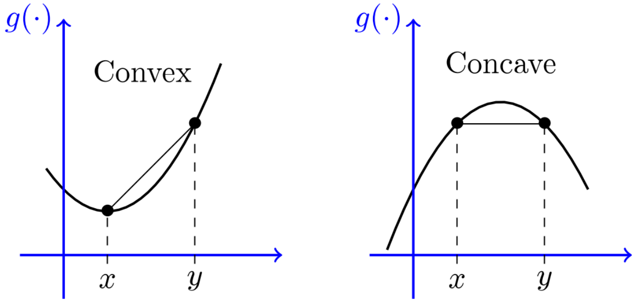
What properties should f have so that the problem is tractable?

What we know from continuous optimization



A function $g : \mathbb{R}^n \to \mathbb{R}$ can be minimized efficiently, if it is convex
A function $g : \mathbb{R}^n \to \mathbb{R}$ can be maximized efficiently, if it is concave

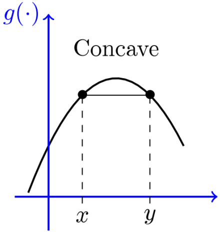
What we know from continuous optimization



A function $g : \mathbb{R}^n \to \mathbb{R}$ can be minimized efficiently, if it is convex
A function $g : \mathbb{R}^n \to \mathbb{R}$ can be maximized efficiently, if it is concave

Can we have analogs in discrete optimization with $f : \{0,1\}^n \to \mathbb{R}$?

From concavity to submodularity



$g'(x) = \lim_{\Delta x \to 0} \frac{g(x+\Delta x) - g(x)}{\Delta x}$
is non-increasing in x

Let $f : 2^N \to \mathbb{R}$ and define

$$\Delta_e f(S) = f(S \cup \{e\}) - f(S)$$

Submodularity: we require that $\Delta_e f(S)$ is non-increasing in S

Submodular functions: definitions

Submodularity: we require that $\Delta_e f(S)$ is non-increasing in S

Submodular functions I

A function $f : 2^N \to \mathbb{R}$ is submodular if for every $A \subseteq B \subseteq N$ and $e \notin B$, it holds that

$$\Delta_e f(A) \geq \Delta_e f(B) \quad \Longleftrightarrow \quad f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B)$$

Submodular functions: definitions

Submodularity: we require that $\Delta_e f(S)$ is non-increasing in S

Submodular functions I

A function $f : 2^N \to \mathbb{R}$ is submodular if for every $A \subseteq B \subseteq N$ and $e \notin B$, it holds that

$$\Delta_e f(A) \geq \Delta_e f(B) \quad \iff \quad f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B)$$

Submodular functions II

A function $f : 2^N \to \mathbb{R}$ is submodular if for every $A, B \subseteq N$, it holds that

$$f(A \cap B) + f(A \cup B) \leq f(A) + f(B)$$

Submodular functions: definitions

Submodularity: we require that $\Delta_e f(S)$ is non-increasing in S

Submodular functions I

A function $f : 2^N \to \mathbb{R}$ is submodular if for every $A \subseteq B \subseteq N$ and $e \notin B$, it holds that

$$\Delta_e f(A) \geq \Delta_e f(B) \quad \iff \quad f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B)$$

Submodular functions II

A function $f : 2^N \to \mathbb{R}$ is submodular if for every $A, B \subseteq N$, it holds that

$$f(A \cap B) + f(A \cup B) \leq f(A) + f(B)$$

Why "submodular"?

Submodular functions: definitions

Submodularity: we require that $\Delta_e f(S)$ is non-increasing in S

Submodular functions I

A function $f : 2^N \to \mathbb{R}$ is submodular if for every $A \subseteq B \subseteq N$ and $e \notin B$, it holds that

$$\Delta_e f(A) \geq \Delta_e f(B) \quad \Longleftrightarrow \quad f(A \cup \{e\}) - f(A) \geq f(B \cup \{e\}) - f(B)$$

Submodular functions II

A function $f : 2^N \to \mathbb{R}$ is submodular if for every $A, B \subseteq N$, it holds that

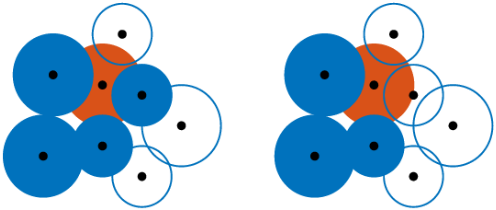
$$f(A \cap B) + f(A \cup B) \leq f(A) + f(B)$$

Why "submodular"? A function $f : 2^N \to \mathbb{R}$ is modular if

$$f(A \cup \{e\}) - f(A) = f(B \cup \{e\}) - f(B) \implies f(A) = \sum_{i \in A} f(\{i\}) + (|A| - 1) f(\emptyset)$$

Submodular functions: examples

Adding additional sensors has diminishing returns

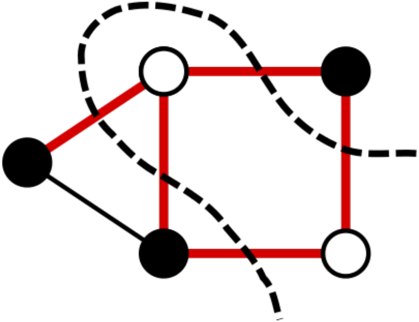


A sensor coverage example

$f : S \mapsto \mathbb{R}$ computes the area that sensors in set S cover

Submodular functions: examples

Including more nodes in a cut has diminishing returns



Cut function $f(S) = u(S, \bar{S}) = \sum_{i \in S, j \in \bar{S}} u_{ij}$

Today

1. Submodular functions

2. Monotone submodular maximization

3. Unconstrained submodular minimization

Monotone submodular maximization & cardinality

Monotone function

A function $f : 2^N \to \mathbb{R}$ is monotone if for every $A \subseteq B \subseteq N$, $f(A) \leq f(B)$.

Monotone submodular maximization & cardinality

Monotone function

A function $f : 2^N \to \mathbb{R}$ is monotone if for every $A \subseteq B \subseteq N$, $f(A) \leq f(B)$.

$$\max\{f(S) : |S| \leq k\} \tag{1}$$

Monotone submodular maximization & cardinality

Monotone function

A function $f : 2^N \to \mathbb{R}$ is monotone if for every $A \subseteq B \subseteq N$, $f(A) \leq f(B)$.

$$\max\{f(S) : |S| \leq k\} \tag{1}$$

Algorithm Greedy algorithm

1: $S \leftarrow \emptyset$
2: **while** $|S| < k$ **do**
3: Pick $i^* \in \text{argmax}_{i \in N \setminus S} f(S \cup \{i\}) - f(S)$
4: $S \leftarrow S \cup \{i^*\}$
5: **end while**

Monotone submodular maximization & cardinality

Monotone function

A function $f : 2^N \to \mathbb{R}$ is monotone if for every $A \subseteq B \subseteq N$, $f(A) \leq f(B)$.

$$\max\{f(S) : |S| \leq k\} \tag{1}$$

Algorithm Greedy algorithm

1: $S \leftarrow \emptyset$
2: **while** $|S| < k$ **do**
3: Pick $i^* \in \text{argmax}_{i \in N \setminus S} f(S \cup \{i\}) - f(S)$
4: $S \leftarrow S \cup \{i^*\}$
5: **end while**

Greedy is "good enough"

Suppose f in (1) is nonnegative, monotone submodular, then the greedy algorithm finds a solution S to (1) such that

$$f(S) \geq (1 - e^{-1}) f(S^*),$$

where S^* is an optimal solution.

Submodular maximization: more settings

Monotone maximization

- (1978) Under intersection of k matroids ($\frac{1}{k+1}$)
- (2004) Under knapsack constraint ($1 - e^{-1}$)

$$\max\{f(S) : \sum_{i \in S} c_i \leq B\}$$

  where c_i's are nonnegative integers
- (2007) Under a matroid constraint ($1 - e^{-1}$)
- ...

Non-monotone maximization

- (2007) Unconstrained ($\frac{1}{2}$)
- (2011) Under cardinality or matroid constraint (e^{-1})
- ...

Today

1. Submodular functions
2. Monotone submodular maximization
3. Unconstrained submodular minimization

From discrete optimization to continuous optimization

$$\min\{f(S) : S \subseteq N\}$$

- $f : \{0,1\}^n \to \mathbb{R}$ is a function whose domain contains 2^n points
- Can we define a continuous extension $f^{\mathrm{E}} : [0,1]^n \to \mathbb{R}$?
- Hopefully
  1. f^{E} coincides with f at $\{0,1\}^n$
  2. f^{E} can be easily evaluated
  3. f^{E} can be efficiently optimized
  4. An optimizer S^* of f can be recovered from an optimizer x^* of f^{E}

From discrete optimization to continuous optimization

$$\min\{f(S) : S \subseteq N\}$$

- $f : \{0,1\}^n \to \mathbb{R}$ is a function whose domain contains 2^n points
- Can we define a continuous extension $f^{\mathrm{E}} : [0,1]^n \to \mathbb{R}$?
- Hopefully
  1. f^{E} coincides with f at $\{0,1\}^n$
  2. f^{E} can be easily evaluated
  3. f^{E} can be efficiently optimized
  4. An optimizer S^* of f can be recovered from an optimizer x^* of f^{E}

Extension of a set function

Given a set function $f : \{0,1\}^n \to \mathbb{R}$, an extension of f to $[0,1]^n$ is a function $f^{\mathrm{E}} : [0,1]^n \to \mathbb{R}$ satisfying $f^{\mathrm{E}}(x) = f(x)$ for $x \in \{0,1\}^n$.

Solve a continuous optimization problem with the extension of f.

Note $f : \{0,1\}^n \to \mathbb{R}$ is defined only at 2^n locations, for $x \in [0,1]^n$:

1. Find coefficients α_i's such that $x = \sum_i \alpha_i \mathbb{1}_{S_i}$
2. Define $f^{\mathrm{E}}(x) = \sum_i \alpha_i f(S_i)$

Continuous extensions

Note $f : \{0,1\}^n \to \mathbb{R}$ is defined only at 2^n locations, for $x \in [0,1]^n$:

1. Find coefficients α_i's such that $x = \sum_i \alpha_i \mathbb{1}_{S_i}$
2. Define $f^{\mathrm{E}}(x) = \sum_i \alpha_i f(S_i)$

- The combinations may not be unique

$$x = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \quad = \quad \frac{1}{2}\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad = \quad \frac{1}{2}\begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Note $f : \{0,1\}^n \to \mathbb{R}$ is defined only at 2^n locations, for $x \in [0,1]^n$:

1. Find coefficients α_i's such that $x = \sum_i \alpha_i \mathbb{1}_{S_i}$
2. Define $f^{\mathrm{E}}(x) = \sum_i \alpha_i f(S_i)$

- The combinations may not be unique

$$x = \begin{bmatrix} \frac{1}{2} \\ \frac{1}{2} \end{bmatrix} \quad = \quad \frac{1}{2}\begin{bmatrix} 1 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad = \quad \frac{1}{2}\begin{bmatrix} 0 \\ 0 \end{bmatrix} + \frac{1}{2}\begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

- If every $S_i \subseteq N$ appears in the sum, then exponentially many terms

The Lovász extension: definition

The Lovász extension: definition

Given $f : \{0,1\}^n \to \mathbb{R}^n$, its Lovász extension $f^{\mathsf{L}} : [0,1]^n \to \mathbb{R}$ is defined by

$$f^{\mathsf{L}}(x) = \mathbb{E}_\theta[f(\{i : x_i \geq \theta\})]$$

where θ is a random variable following a uniform distribution over $[0,1]$.

The Lovász extension: definition

The Lovász extension: definition

Given $f : \{0,1\}^n \to \mathbb{R}^n$, its Lovász extension $f^{\mathsf{L}} : [0,1]^n \to \mathbb{R}$ is defined by

$$f^{\mathsf{L}}(x) = \mathbb{E}_\theta[f(\{i : x_i \geq \theta\})]$$

where θ is a random variable following a uniform distribution over $[0,1]$.

Let $\pi : \{1,\ldots,n\} \to \{1,\ldots,n\}$ be a permutation of x such that

$$x_{\pi(0)} = 1 \geq x_{\pi(1)} \geq \cdots \geq x_{\pi(n)} \geq 0 = x_{\pi(n+1)}$$

- When $x_{\pi(i+1)} < \theta \leq x_{\pi(i)}$, we have $\{i : x_i \geq \theta\} = \{\pi(1),\ldots,\pi(i)\}$
- The expectation defining $f^{\mathsf{L}}(x)$ is then a sum of $n+1$ terms
- f^{L} and f coincide at $x \in \{0,1\}^n$

The Lovász extension as a continuous extension

Note $f : \{0,1\}^n \to \mathbb{R}$ is defined only at 2^n locations, for $x \in [0,1]^n$:

1. Find coefficients α_i's such that $x = \sum_i \alpha_i \mathbb{1}_{S_i}$
2. Define $f^{\mathrm{E}}(x) = \sum_i \alpha_i f(S_i)$

Let $x \in [0,1]^n$, suppose $x_1 \leq x_2 \leq \cdots \leq x_n$, note

$$x = x_1 \begin{bmatrix} 1 \\ 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{bmatrix} + (x_2 - x_1) \begin{bmatrix} 0 \\ 1 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{bmatrix} + (x_3 - x_2) \begin{bmatrix} 0 \\ 0 \\ 1 \\ \vdots \\ \vdots \\ 1 \end{bmatrix} + \cdots + (x_n - x_{n-1}) \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 1 \end{bmatrix} + (1 - x_n) \begin{bmatrix} 0 \\ 0 \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix}$$

Thus,

$$f^{\mathrm{L}}(x) = x_1 f(S) + (x_2 - x_1) f(S \setminus \{1\}) + \cdots + (x_n - x_{n-1}) f(\{n\}) + (1 - x_n) f(\emptyset)$$

- $f(S) = \min\{1, |S|\}$, that is $f(\emptyset) = 0$ and $f(S) = 1$ elsewhere

The Lovász extension: examples

- $f(S) = \min\{1, |S|\}$, that is $f(\emptyset) = 0$ and $f(S) = 1$ elsewhere

$$f^{\mathsf{L}}(x) = \mathbb{E}_\theta[f(\{i : x_i \geq \theta\})] = \max_i\{x_i\}$$

The Lovász extension: examples

- $f(S) = \min\{1, |S|\}$, that is $f(\emptyset) = 0$ and $f(S) = 1$ elsewhere

$$f^{\mathrm{L}}(x) = \mathbb{E}_\theta[f(\{i : x_i \geq \theta\})] = \max_i\{x_i\}$$

- Let $G = (N, E)$ be an undirected graph, for $A \subseteq V$, define

$$f(A) = |\{(i,j) \in E \mid i \in A, j \in \bar{A}\}|$$

The Lovász extension: examples

- $f(S) = \min\{1, |S|\}$, that is $f(\emptyset) = 0$ and $f(S) = 1$ elsewhere

$$f^{\mathsf{L}}(x) = \mathbb{E}_\theta[f(\{i : x_i \geq \theta\})] = \max_i \{x_i\}$$

- Let $G = (N, E)$ be an undirected graph, for $A \subseteq V$, define

$$f(A) = |\{(i,j) \in E \mid i \in A, j \in \bar{A}\}|$$

Then the Lovász extension is given by

$$f^{\mathsf{L}}(x) = \sum_{(i,j)\in E} |x_i - x_j|$$

Submodular minimization: setups

$$\min\{f(S) : S \subseteq N\}$$

- $f : \{0,1\}^n \to \mathbb{R}$ is submodular
- No monotonicity is assumed
- No constraint is imposed

Key idea: Lovász extension

$$\begin{array}{c}\text{Submodular function:}\\ f : \{0,1\}^n \to \mathbb{R}\\ \Downarrow \\ \text{Convex continuous function:}\\ f^{\mathrm{L}} : [0,1]^n \to \mathbb{R}\end{array}$$

László Lovász (1948-)

- If f^{L} is convex, then the extension can be solved efficiently
- A minimizer of f^{L} can be converted into a minimizer of $f(S)$

The Lovász extension: why useful

The Lovász extension: definition

Given $f : \{0,1\}^n \to \mathbb{R}^n$, its Lovász extension $f^{\mathrm{L}} : [0,1]^n \to \mathbb{R}$ is defined by

$$\begin{aligned} f^{\mathrm{L}}(x) &= \sum_{i=0}^{n} \lambda_i(x) f(S_i(x)) \\ &= f(\emptyset) + \sum_{i=1}^{n} x_{\pi(i)}\big(f(S_i(x)) - f(S_{i-1}(x))\big) \end{aligned}$$

1. $\lambda_i(x) = x_{\pi(i)} - x_{\pi(i+1)}$
2. $S_{i+1} = S_i \cup \{\pi(i+1)\}$ with $S_0 = \emptyset$

The Lovász extension is a piece-wise linear function of x!

The Lovász extension: why useful

The Lovász extension: definition

Given $f : \{0,1\}^n \to \mathbb{R}^n$, its Lovász extension $f^{\mathsf{L}} : [0,1]^n \to \mathbb{R}$ is defined by

$$\begin{aligned} f^{\mathsf{L}}(x) &= \sum_{i=0}^{n} \lambda_i(x) f(S_i(x)) \\ &= f(\emptyset) + \sum_{i=1}^{n} x_{\pi(i)}(f(S_i(x)) - f(S_{i-1}(x))) \end{aligned}$$

1. $\lambda_i(x) = x_{\pi(i)} - x_{\pi(i+1)}$
2. $S_{i+1} = S_i \cup \{\pi(i+1)\}$ with $S_0 = \emptyset$

The Lovász extension is a piece-wise linear function of x!

Convexity and submodularity

A function $f : \{0,1\}^n \to \mathbb{R}^n$ is submodular if and only if its Lovász extension $f^{\mathsf{L}} : [0,1]^n \to \mathbb{R}$ is convex.

The Lovász extension: why useful

Convexity and submodularity

A function $f : \{0,1\}^n \to \mathbb{R}^n$ is submodular if and only if its Lovász extension $f^{\mathsf{L}} : [0,1]^n \to \mathbb{R}$ is convex.

- We can recover an optimal set S^* from a minimizer x^* of $f^{\mathsf{L}}(x)$

$$f^{\mathsf{L}}(x^*) = \sum_{i=0}^{n} \lambda_i(x^*) f(S_i(x^*))$$

  where $\sum_{i=0}^{n} \lambda_i(x^*) = 1$

We did not discuss

- Why is $f^{\mathsf{L}}(x)$ convex when f is submodular?
- How to solve the convex problem?

Upcoming

Week 1-8 (AU4606 & AI4702):

- Introduction (1 lecture)
- Preparations (3 lectures)
  - basics of graph theory
  - algorithm complexity and data structure
  - graph search algorithm
- Shortest path problems (3 lectures)
- Maximum flow problems (5 lectures)
- Minimum cost flow problems (3 lectures)
- Introduction to multi-agent systems (1 lecture)
- Introduction to cloud networks (1 lecture)

Week 9-16 (AU4606):

- Simplex and network simplex methods (1 lecture)
- Global minimum cut problems (1.5 lectures)
- Minimum spanning tree problems (1.5 lectures)
- Submodular function optimization (2 lectures)
- Optimal assignments and matching (2 lectures)