Optimization Models and Formulations I

Yinyu Ye
Department of Management Science and Engineering
Stanford University
Stanford, CA 94305, U.S.A.

https://canvas.stanford.edu/courses/179677

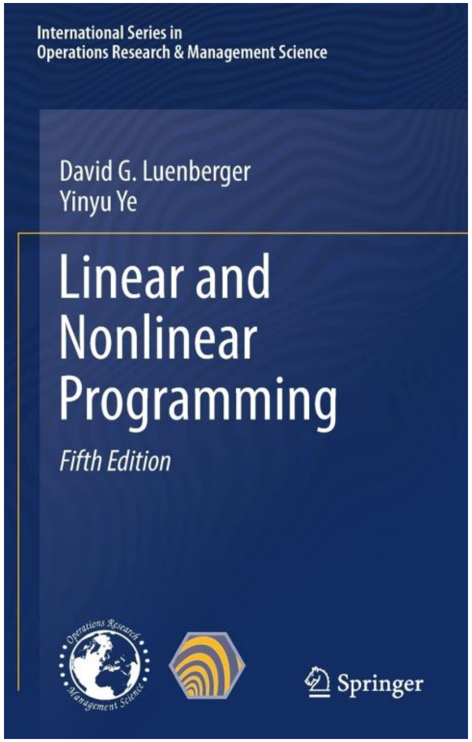
Read Chapter 1.1, 1.2, 2.1, 2.2, Appendices A, B&D in Text-Book (hard copies would be available in the Book Store)

1st Day Questions

- My CA team: Xuhui, Chunlin, J.T., Eva, Celia
- Websites: Canvas

  https://canvas.stanford.edu/courses/142617
- 5 Homework assignments (count the best 4),

  One in-class Midterm, one team project
  - 40%*H + 30%*M+ 30%*P
  - No difference on taking 3 or 4 units
- No formula for cutoff between A/B etc.
- The more fun we all have, the more A's we will give out.
- Textbook: Linear and Nonlinear Programming (LY 5th edition)
- This is 111/211; take 111X/211X for more advanced levels
- The software use will help: Solvers in Matlab, R, Python or other public free software.
- Form a "diversified" study group
- Selected Friday's problem sessions (will be taped)
- Students with OAE, extra time for the exam



Mathematical Optimization Models

- Often consider the common quantitative model of data/decision/management science & engineering:

  - Maximize or Minimize $f(\boldsymbol{x})$
    for all $\boldsymbol{x} \in$ some set X

- Decision variables $\boldsymbol{x}$, Objective function $f(\boldsymbol{x})$, Constraint set X

- Applications in:
  - Applied Science, Engineering, Economics, Finance, Medicine, Statistics, Business
  - General Decision and Policy Making

- The famous Eighteenth Century Swiss mathematician and physicist Leonhard Euler (1707-1783) proclaimed that "…nothing at all takes place in the Universe in which some rule of maximum or minimum does not appear."

The Prototypical Optimization Problems

Max (or Max): $f(\mathbf{x})$

s.t. :

$$h_1(\mathbf{x}) = 0$$

$$\ldots$$

$$h_m(\mathbf{x}) = 0$$

$$g_1(\mathbf{x}) \leq 0$$

$$\ldots$$

$$g_r(\mathbf{x}) \leq 0$$

The Function could be: x_1+2x_2, $x^2+2xy+2y^2$, $x\ln(x)+e^y$, $|x|+\max\{x,y\}$, etc

Linear Programming/Optimization: all functions are linear/affine

- Decision Variables $\mathbf{x} \in R^n$, yet to be decided, form a **solution; Objective function** $f(x)$
- Constraints: **equality, inequality; feasible solutions**: meet all constraints
- Optimal solution x^*: **feasible** and achieve the **best** possible objective value $f(\mathbf{x}^*)$

3 Main Categories Covered in this Course

- **Linear Optimization** (Programming)
  - Problem Formulation, Optimality Conditions
  - Search Algorithms, e.g., Simplex and Interior-Point Algorithms
- **Unconstrained Nonlinear Optimization**
  - Problem Formulation, Optimality Conditions
  - 1^{st} order methods, gradient method; 2^{nd} order methods, Newton
- **Constrained Nonlinear Optimization**
  - Problem Formulation, Lagrangian Functions, Optimality Conditions
  - 1^{st} order, gradient projection, sequential LP,; 2^{nd} order, sequential Newton
  - Duality Theories,

Other Classifications:
Quadratic, Convex, Integer, Mixed-Integer, Binary, etc.

Issues in Optimization

- **How to formulate a real-life problem**
  – Three steps: variables, objective, constraints

- **How to recognize a solution being optimal?**
  – Easy to check

- **How to measure algorithm effciency?**
  – Convergence speed
  – Local Convergence speed

- **Insight more than just the solution**?
  – Solution structure properties
  – Sensitivity analysis
  – Alternate formulations

What do you learn?

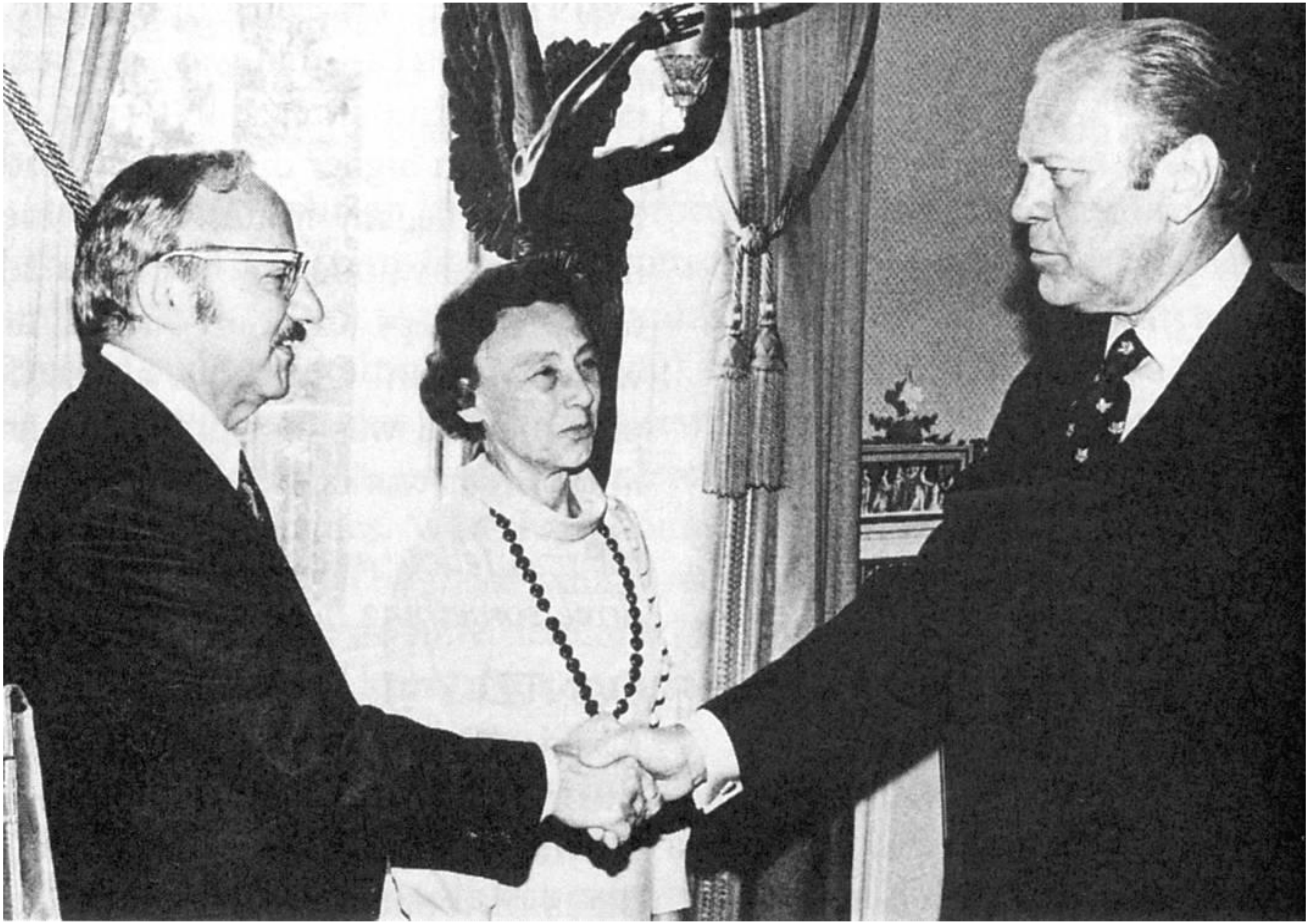
- Models – the Art: intuition and common sense
  - How formulate real problems using quantitative models
- Theory – the Science: theorems, geometries and universal rules
  - Necessary and Sufficient Conditions that must be true for the optimality of different classes of problems.
- Algorithms – the Engineering: algorithms, methodologies and software tools
  - How we apply the theory to robustly and efficiently solve problems and gain insight beyond the solution.
- Applications – AI, Machine Learning and Data Science
  - Logistic Regression, SVM, the Wasserstain barycenter, Reinforced learning/MDP, Information market,…

Start with Linear Programming

- Why do we study LP's
  - Not just because solving non-linear problems are difficult
  - But because real-world problems are often formulated as linear equations and inequalities
    - Either because they indeed are linear
    - Or because it is unclear how to represent them and linear is an intuitive compromise
  - A stepping stone for solving more complicated nonlinear optimization problems, which you would see later.

LP, Nobel Prize,…

… and National Medal of Science



Optimization Problem Modeling/Formulation

Modeling Process

- Understand the problem
- Collect relevant information and data
- Identify and define the decision variables
  - The role of variables
  - How they related to data
- formulate the objective function
  - Logic relation to Data and Decision Variables
- Isolate and formulate the constraints
  - From physical conditions, regularity limits, common senses

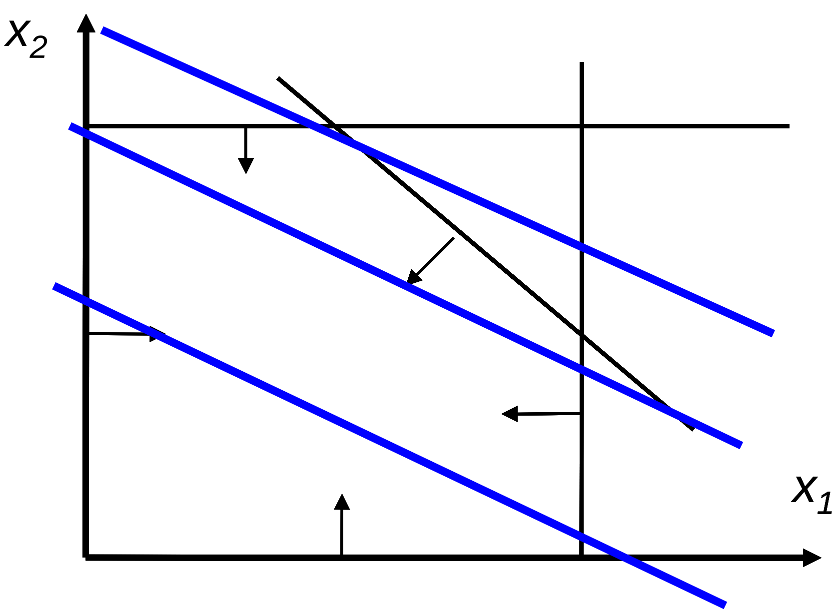
All based on Science, Arts, and Intuitions….

LP Example 1: Resource Allocation/Production Management

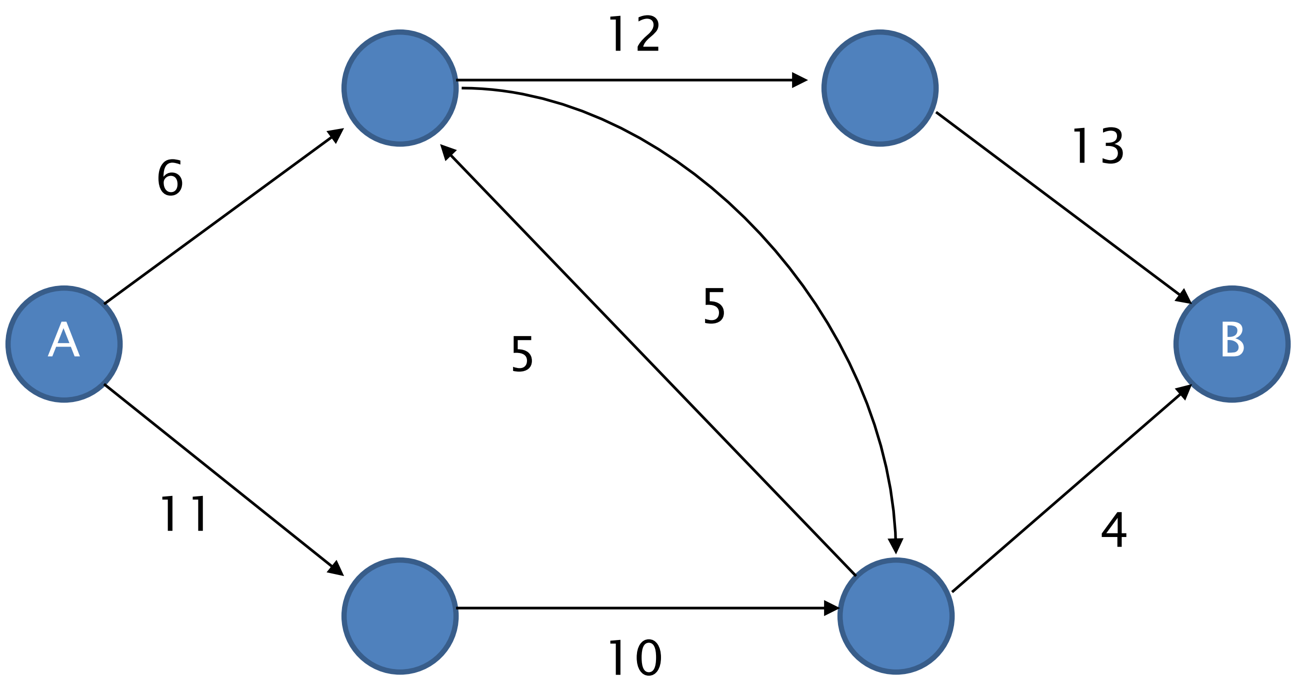
The Wyndor Glass Co. is a producer of high-quality glass products. It has three plants. Aluminum frames and hardware are made in Plant 1, wood frames are made in Plant 2, and Plant 3 is used to produce glass and assemble the products. Wyndor produces two products which require the resources of the three plants as follows:

Plant	Aluminum	Wood	Resources
1	1	0	100
2	0	2	200
3	1	1	150
Unit Profit	$1000	$2000	

$$
\begin{array}{ll}
\max & x_1 + 2x_2 \\
\text{s.t.} & x_1 \le 1, \\
& 2x_2 \le 2, \\
& x_1 + x_2 \le 1.5, \\
& x_1, x_2 \ge 0
\end{array}
$$

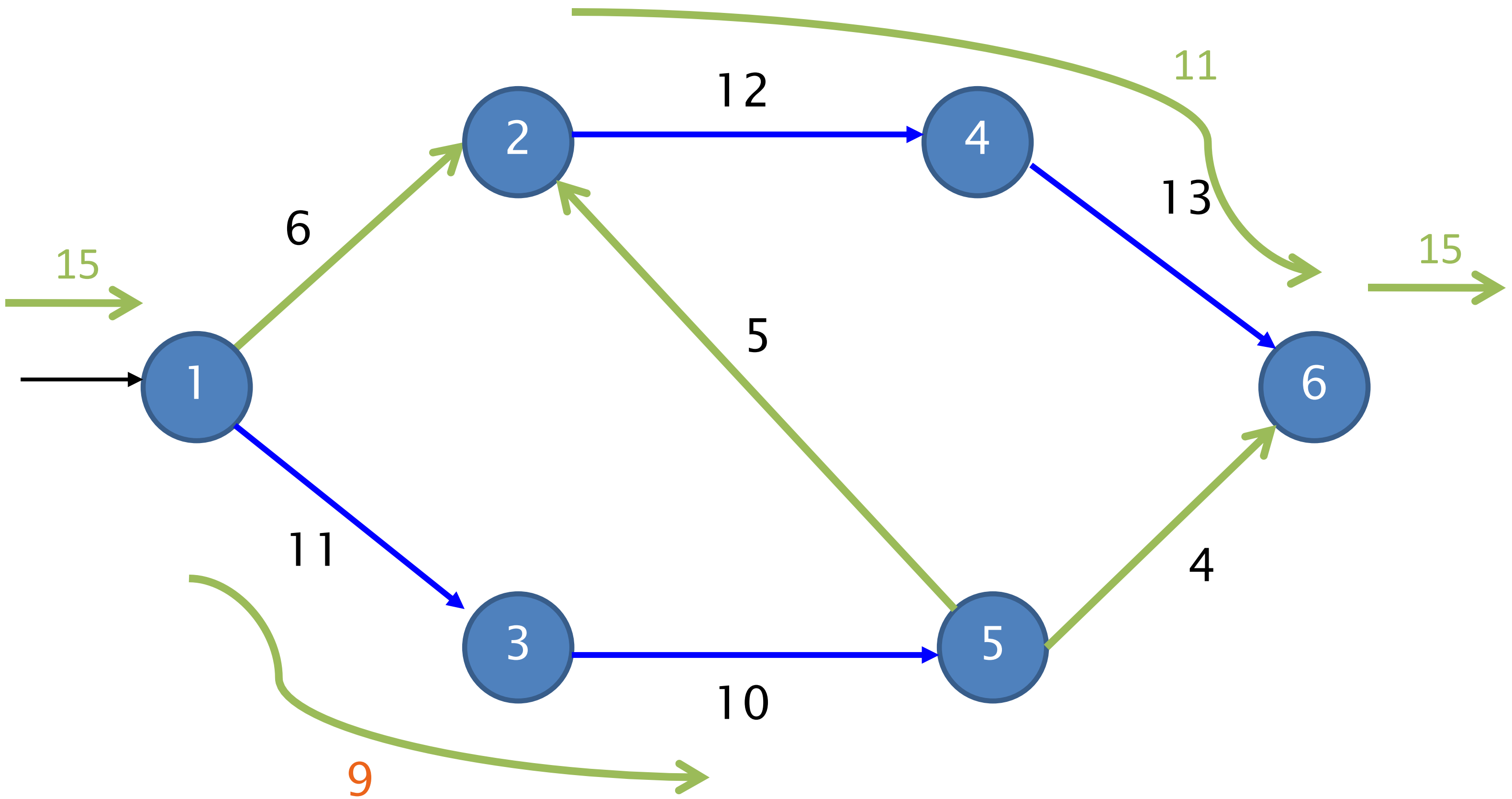


LP Example 2: Maximum Flow



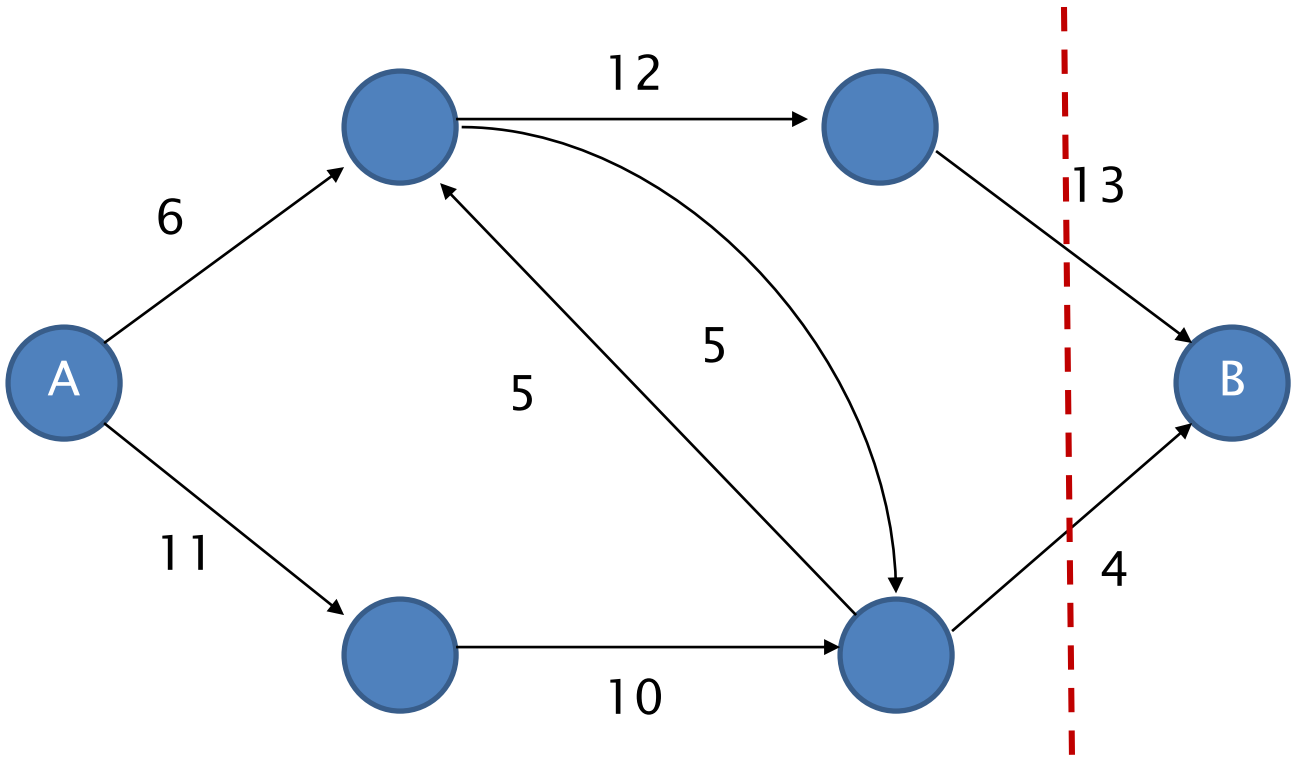
How much flow can travel from A to B, given that each of the directed connecting routes have flow limits/capacities?

Maximum Flow by Inspection



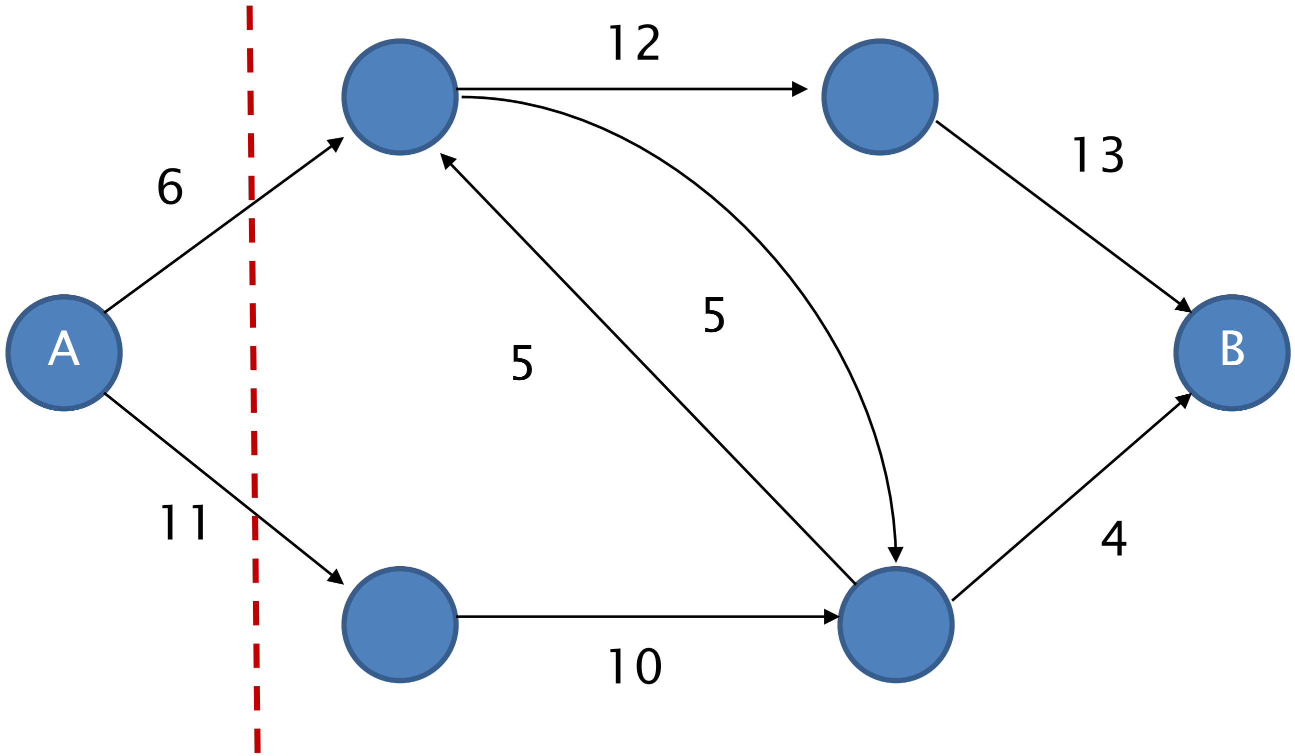
How to certify that 15 is maximal?

Cut in Maximum Flow I



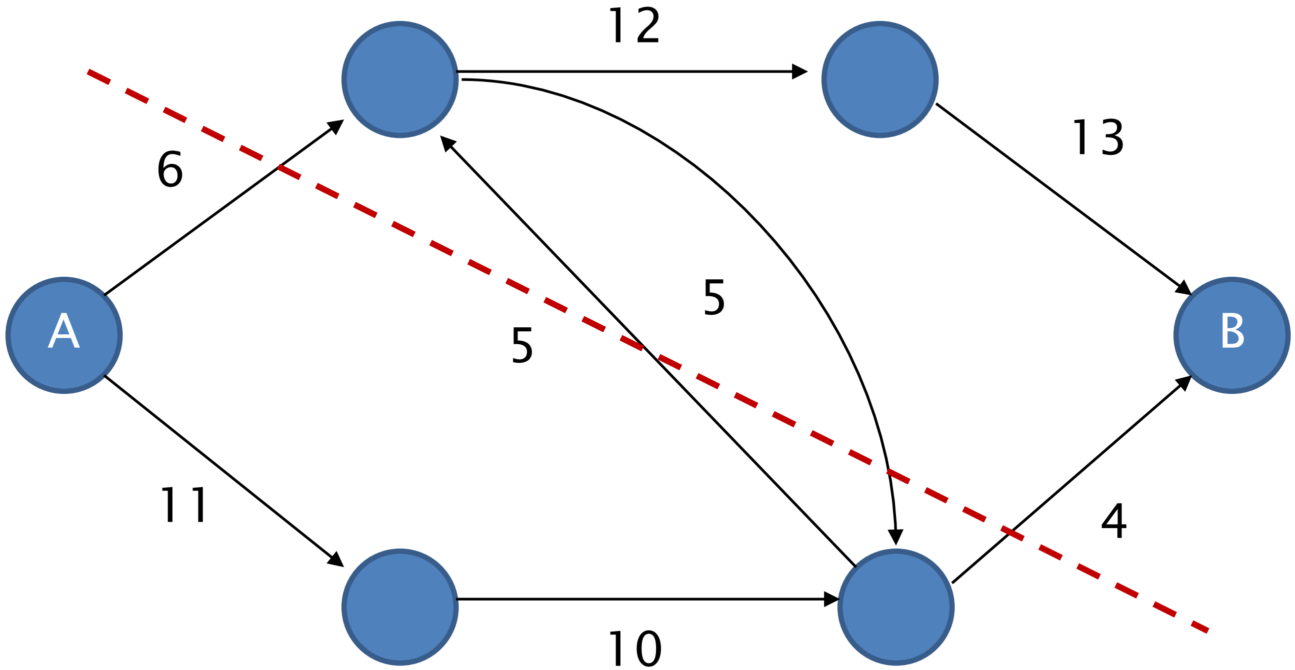
Cut value from Source site to Sink site=17

Cut in Maximum Flow II



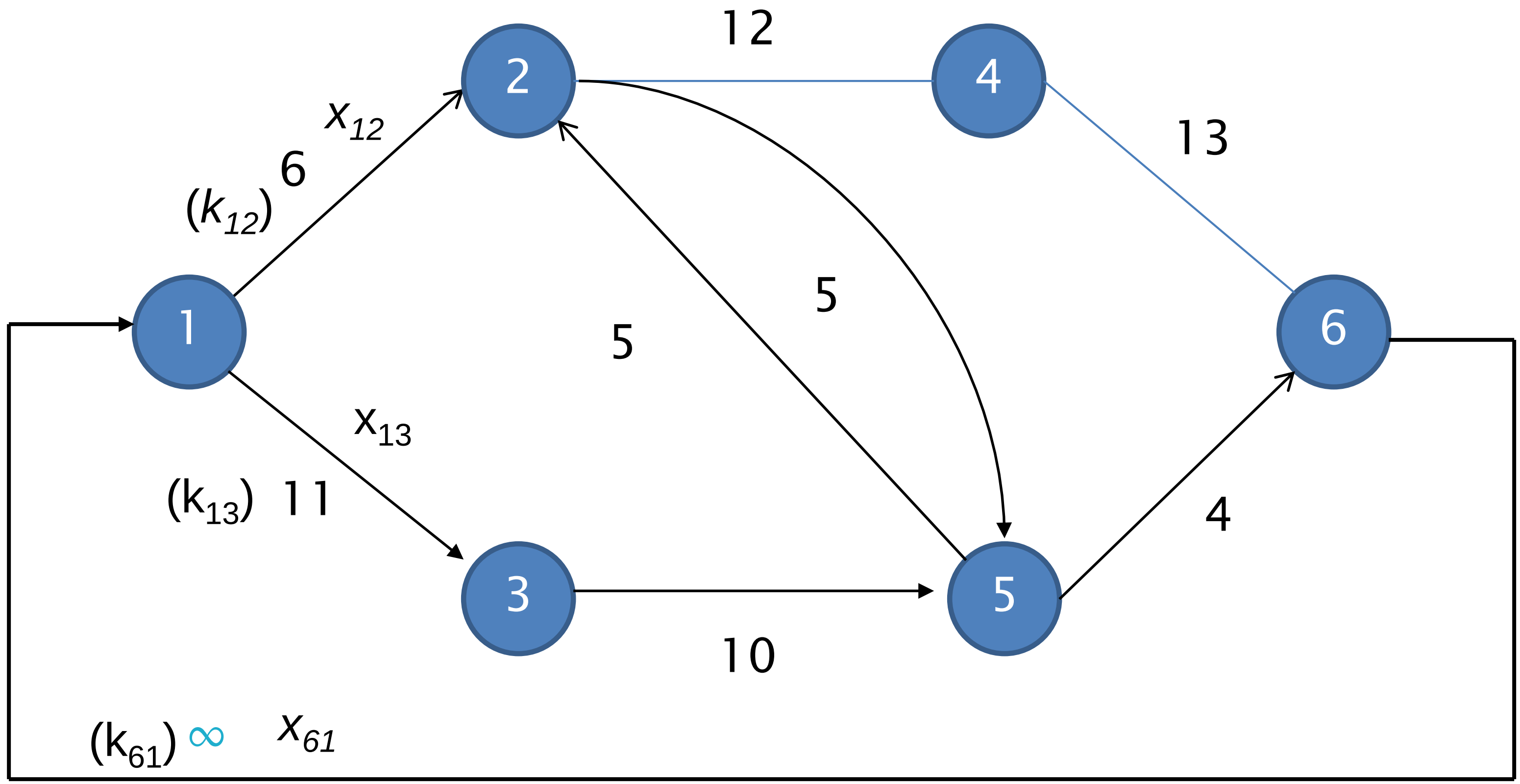
Cut value from Source site to Sink site=17

Cut in Maximum Flow III



Cut value from Source site to Sink site=15

Data points classification application in Machine Learning and Data Science



$$\max \quad x_{61}$$

$$\text{s.t.} \quad \sum_k x_{ki} = \sum_j x_{ij} \quad \forall\, i = 1,2,3,4,5,6$$

$$0 \le x_{ij} \le k_{ij} \quad \forall\, i, j = 1,2,3,4,5,6$$

inflow

outflow

LP Example 3: Electric Vehicle Charging Schedule and Storage Control

	Period 1	Period 2	Period 3	Period 4	Period 5
Price ($)	**1.25 (c_1)**	**1.35 (c_2)**	**1.25 (c_3)**	**1.10 (c_4)**	**1.05 (c_5)**
Demand (kw)	**60 (d_1)**	**110 (d_2)**	**100 (d_3)**	**40 (d_4)**	**0 (d_5)**
Charging (kw)	x_1	x_2	x_3	x_4	x_5
Inventory (I_0)	I_1	I_2	I_3	I_4	I_5

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{5} c_i x_i \\
\text{s.t.} \quad & I_{i-1} + x_i - d_i = I_i, \quad \forall\, i = 1,2,3,4,5 \\
& I_{i-1} + x_i \le K, \quad \forall\, i = 1,2,3,4,5 \\
& x_i \ge 0, I_i \ge 0, \quad \forall\, i.
\end{aligned}
$$

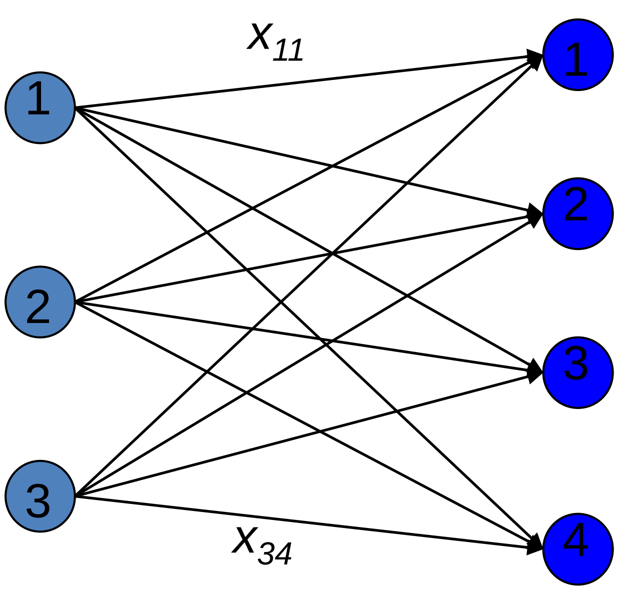
When Discharge is Allowed

	Period 1	Period 2	Period 3	Period 4	Period 5
Price ($)	1.25 (c_1)	1.35 (c_2)	1.25 (c_3)	1.10 (c_4)	1.05 (c_5)
Demand (kw)	60 (d_1)	110 (d_2)	100 (d_3)	40 (d_4)	0 (d_5)
Charging (kw)	x_1	x_2	x_3	x_4	x_5
Inventory (I_0)	I_1	I_2	I_3	I_4	I_5

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{5} c_i x_i \\
\text{s.t.} \quad & I_{i-1} + x_i - d_i = I_i, \quad \forall\, i = 1,2,3,4,5 \\
& I_{i-1} + x_i \le K, \quad \forall\, i = 1,2,3,4,5 \\
& I_i \ge 0, \quad \forall\, i.
\end{aligned}
$$

LP Example 4: Transportation and Assignment

	Retailer 1	Retailer 2	Retailer 3	Retailer 4	SUPPLY
Warehouse 1	12 (c_{11})	13	4	6	500 (s_1)
Warehouse 2	6	4	10	11	700 (s_2)
Warehouse 3	10	9	12	14 (c_{34})	800 (s_3)
DEMAND	400 (d_1)	900 (d_2)	200 (d_3)	500 (d_4)	2000 (s_4)



$$\min \quad \sum_{i=1}^{3}\sum_{j=1}^{4} c_{ij}x_{ij}$$

$$\text{s.t.} \quad \sum_{j=1}^{4} x_{ij} = s_i, \quad \forall\, i = 1,2,3$$

$$\sum_{i=1}^{3} x_{ij} = d_j, \quad \forall\, j = 1,2,3,4$$

$$x_{ij} \geq 0, \quad \forall\, i, j$$

Abstract Model

Inventory Planning: ***s*** is part of the decision vars.

Machine Learning: The Wassestein Barycenter Problem I

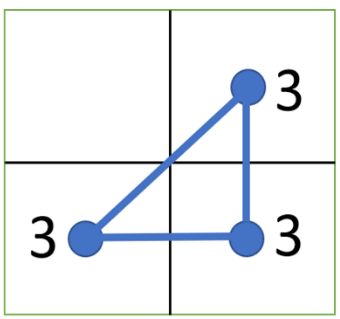
The minimal transportation cost in Data Science is called the Wasserstein distance between a supply distribution and a demand distribution.

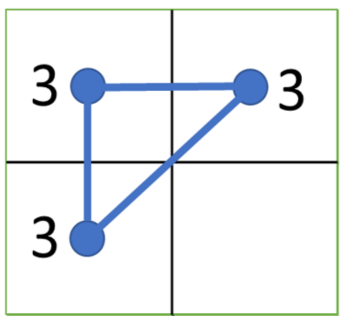
The **Wasserstein Barycenter Problem** is to find a distribution/points such that the sum of its Wasserstein distances to each of a set of distributions/points would be minimized

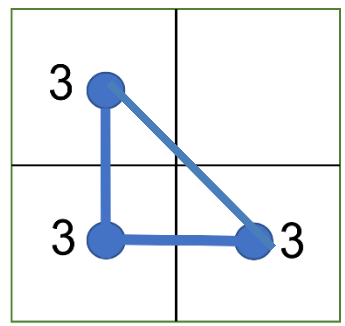
$$\min_s \sum_k \textbf{WD}(s, d_k) \textbf{ s.t. total mass constraint}$$

WD(s, d_k)=

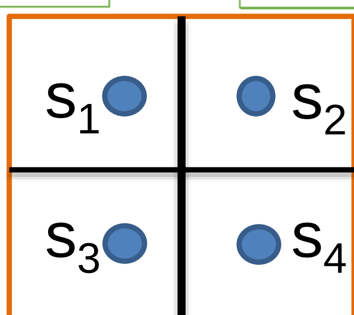
$$\begin{aligned} \min \quad & \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}x_{ij} \\ \text{s.t.} \quad & \sum_{j=1}^{N} x_{ij} = s_i, \quad \forall\, i = 1,\dots,N \\ & \sum_{i=1}^{N} x_{ij} = d_{kj}, \quad \forall\, j = 1,\dots,N \\ & x_{ij} \ge 0, \qquad \forall\, i,j \end{aligned}$$







Three possible demand distribution scenario of 4 cities



Constraints:

$s_1+s_2+s_3+s_4=9$

$(s_1,s_2,s_3,s_4)>=0$

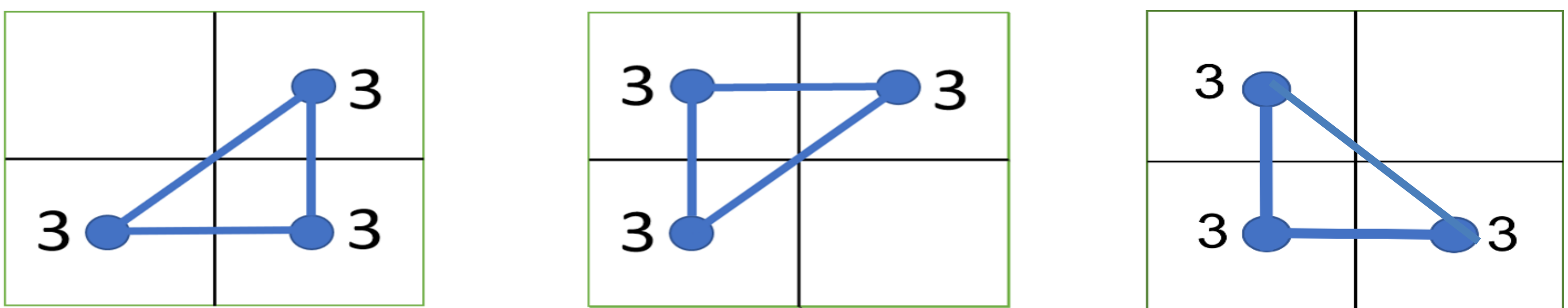
$$C=\begin{pmatrix} 0 & 1 & 1 & 2 \\ 1 & 0 & 2 & 1 \\ 1 & 2 & 0 & 1 \\ 2 & 1 & 1 & 0 \end{pmatrix}$$

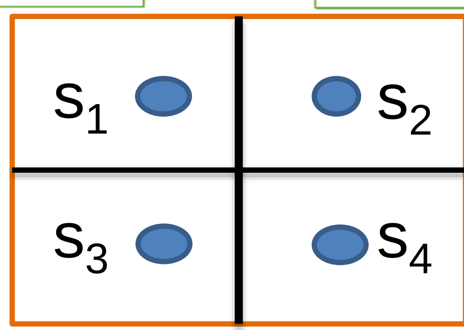
Machine Learning: The Wassestein Barycenter Problem II

$$
\begin{aligned}
\min \quad & \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}x_{1ij} \\
\text{s.t.} \quad & \sum_{j=1}^{N} x_{1ij} - s_i = 0, \forall\ i = 1, \dots, N \\
& \sum_{i=1}^{N} x_{1ij} = d_{1j}, \quad \forall\ j = 1, \dots, N \\
& x_{1ij} \geq 0, \qquad \forall\ i, j
\end{aligned}
$$

$$
\begin{aligned}
+ \quad & \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}x_{2ij} \\
& \sum_{j=1}^{N} x_{2ij} - s_i = 0, \forall\ i = 1, \dots, N \\
& \sum_{i=1}^{N} x_{2ij} = d_{2j}, \quad \forall\ j = 1, \dots, N \\
& x_{2ij} \geq 0, \qquad \forall\ i, j
\end{aligned}
$$

$$
\begin{aligned}
+ \quad & \sum_{i=1}^{N}\sum_{j=1}^{N} c_{ij}x_{3ij} \\
& \sum_{j=1}^{N} x_{3ij} - s_i = 0, \quad \forall\ i = 1, \dots, N \\
& \sum_{i=1}^{N} x_{3ij} = d_{3j}, \quad \forall\ j = 1, \dots, N \\
& x_{3ij} \geq 0, \qquad \forall\ i, j
\end{aligned}
$$



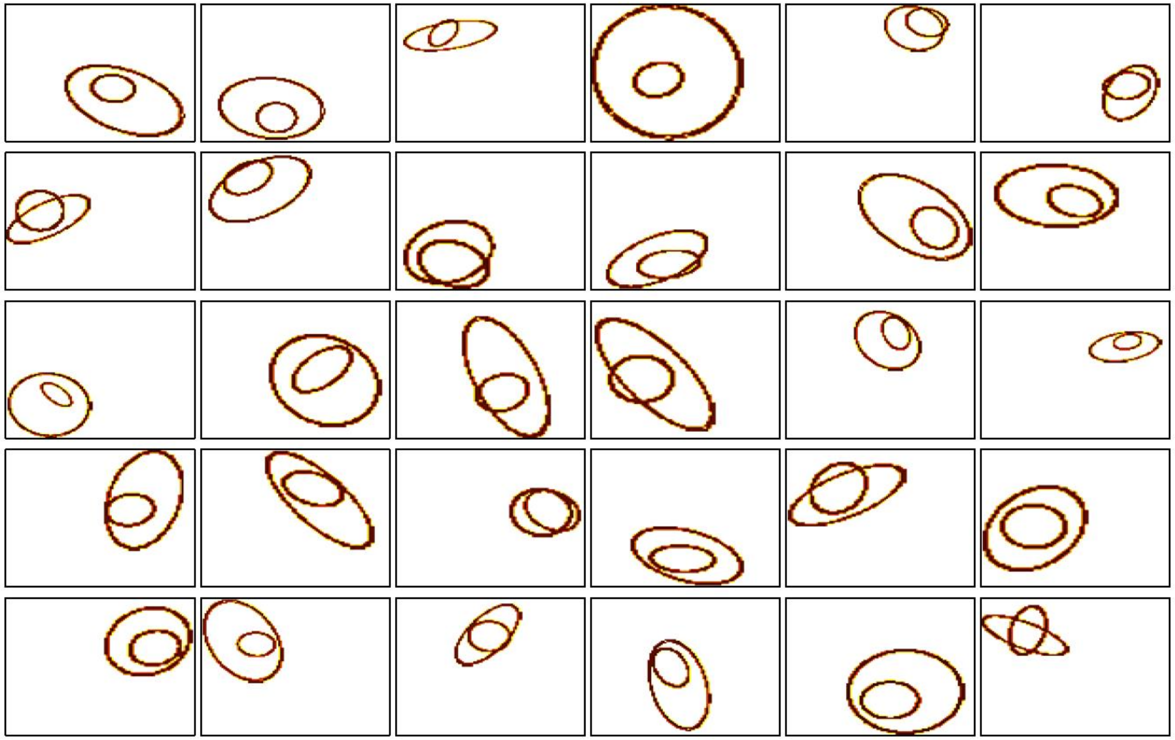


Constraints:

$s_1+s_2+s_3+s_4=9$

$(s_1,s_2,s_3,s_4)>=0$

Machine Learning: The Wassestein Barycenter Problem III



What is the best ``mean or consensus'' image from a set of images (pixel distributions)?

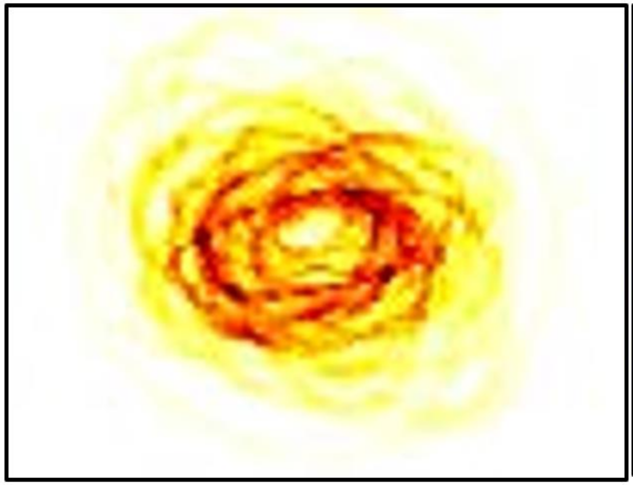
- Simple average
- Simple average after re-centering
- The Wasserstein Barycenter of the set of images (self re-center and rotation)

Machine Learning: The Wassestein Barycenter Problem IV

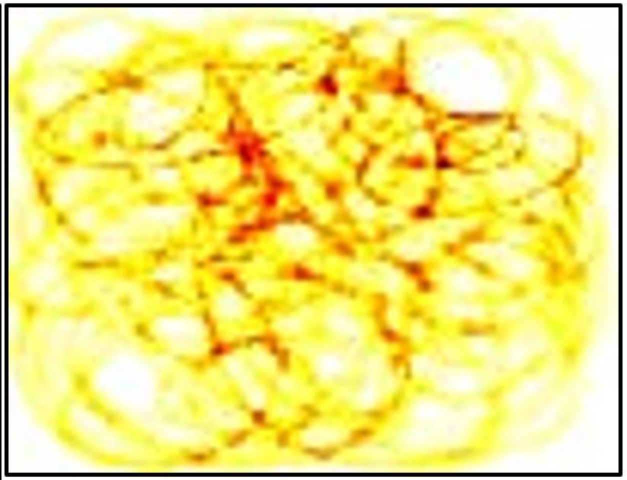
The simple avarage of n points is

$$\mathbf{s}=(\textstyle\sum_k \mathbf{d}_k)/n \quad \text{or} \quad \min_{\mathbf{s}} \textstyle\sum_{\mathbf{k}} (\|\mathbf{s}-\mathbf{d}_k\|_2)^2$$

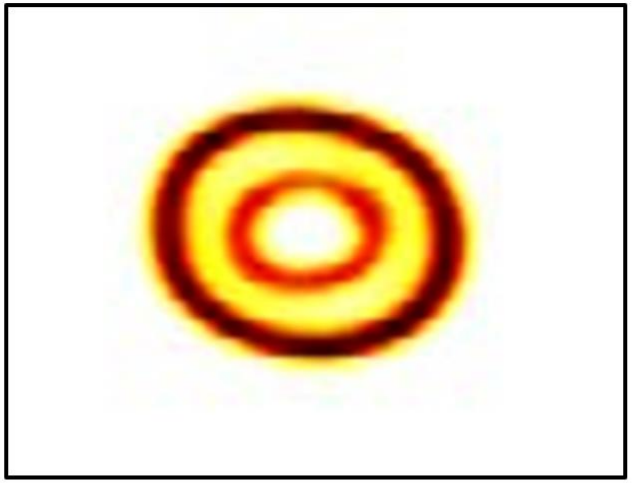
The **Wasserstein Barycenter Problem** is to find a distribution/points such that the sum of its Wasserstein distances to each of a set of distributions/points would be minimized (self re-center and rotation).



Simple average after re-centering



Simple average



the Barycenter image