**Operations Research**

**Lecture Notes**

Prepared & edited by

**Tanmoy Das**

Industrial Engineer & Jr. Data Scientist

Reference materials can be found at

<https://github.com/tanmoyie/Operations-Research>

<https://kaggle.com/tanmoyie>

[www.linkedin.com/in/tanmoyie](http://www.linkedin.com/in/tanmoyie)

**How to read this document**

This document is a reference material along with the topics covered in class of Operations Research (taught by Tanmoy Das). It is agreed that there are some other chapters which are also crucial for the theory course of Operations Research. However, in the class environment, only the following chapters are expected to be covered (Intro to LP, Graphical Solution, Simplex, Dual, Transportation, Machine Learning, Network Optimization, Integer Programming, Game Theory & Queuing Model[[1]](#footnote-1)).

Given that this document is not comprehensive, readers would find relevant materials (including Python codes, scan copies, PDF format, of theoretical and mathematical solutions of the problem discussed from other books) in the following Github repository [***github.com/tanmoyie/Operations-Research***](https://github.com/tanmoyie/Operations-Research)***.*** Download all the pdf, py & other files from the repository to follow accordingly.

There are about fifteen (15) Python projects related to Operations Research (e.g. Travelling Salesman Problem in real world) are covered in this document. YouTube videos related to the explanations of abstruse contents in Operations Research, which involves Python Programming, can be found in

<https://www.youtube.com/playlist?list=PLHyZ7Tamw-fevmrx2V3U13hPDDlUSBbi7>

Some additional Python Projects would be obtained from <https://www.linkedin.com/pulse/python-industrial-engineering-datacamp-level-3-tanmoy-das/>. More contents & YouTube videos will be added shortly. Follow the channel to get more update

www.youtube.com/channel/UC0yUOupBXybIfQ2x7uM6kzg

**Reference Book:**

1. Operations Research (2nd edition) by R. Panneerselvam (Pupils might find this book convenient)
2. Introduction to Operations Research (7th edition) by Lieberman (commonplace textbook)
3. Introduction to OR - deterministic model by Juraj Stacho[[2]](#footnote-2) (an invaluable compendium)

Table of Contents

[Introduction to Operations Research 5](#_Toc516011044)

[Linear Programming 6](#_Toc516011045)

[Math Problem 6](#_Toc516011046)

[Graphical Solution 7](#_Toc516011047)

[Simplex & Dual 8](#_Toc516011048)

[Simplex Method 9](#_Toc516011049)

[Tabular & Big M Solution 9](#_Toc516011050)

[Unbounded & Infeasible Solution 9](#_Toc516011051)

[Revised Simplex 9](#_Toc516011052)

[Dual Problem 9](#_Toc516011053)

[Python Projects on Simplex, Revised Simplex Optimization 10](#_Toc516011054)

[Transportation & Assignment 13](#_Toc516011055)

[NorthWest Corner Method 14](#_Toc516011056)

[Assignment Problem 14](#_Toc516011057)

[Python Projects on Optimization in Transportation & Assignment 15](#_Toc516011058)

[Optimization in Machine Learning/ Data Science 23](#_Toc516011059)

[Linear Regression 24](#_Toc516011060)

[Robust Regression 24](#_Toc516011061)

[Support Vector Machine 24](#_Toc516011062)

[Python Projects on Machine Learning Optimization 25](#_Toc516011063)

[Network Optimization 28](#_Toc516011064)

[Min Cost 29](#_Toc516011065)

[Max Flow 29](#_Toc516011066)

[Minimum Spanning Tree 29](#_Toc516011067)

[Python Projects on Network Optimization 30](#_Toc516011068)

[Integer Programming 38](#_Toc516011069)

**Python projects in Optimization**

[Python Code of Optimization Project 1: Revised Simplex 10](#_Toc516011070)

[Python Code of Optimization Project 2: Shelf Space Optimization 10](#_Toc516011071)

[Python Code of Optimization Project 3: Product Distribution Problem for the PuLP Modeller 16](#_Toc516011072)

[Python Code of Optimization Project 4: Travelling Salesman Problem 20](#_Toc516011073)

[Python Code of Optimization Project 5: Linear Regression (plot\_ols.py) 25](#_Toc516011074)

[Python Code of Optimization Project 6: Robust Regression compared to Linear Regression (plot\_ransac.py) 26](#_Toc516011075)

[Python Code of Optimization Project 7: Min Cost Flow problem 30](#_Toc516011076)

[Python Code of Optimization Project 8: Max Flow 31](#_Toc516011077)

[Python Code of Optimization Project 9: Airlines Network Optimization 32](#_Toc516011078)

# Introduction to Operations Research

****

**Operations research** is a discipline that deals with the application of advanced analytical methods to help make better decisions. **Operational Research always try to find the best and optimal solution to the problem.** For this purpose, objectives of the organization are defined and analyzed. These objectives are then used as the basis to compare the alternative courses of action.

**Optimization Approach:**

**1.** Define the problem of interest and gather relevant data.

**2.** Formulate a mathematical model to represent the problem.

**3.** Develop a computer-based procedure for deriving solutions to the problem from the model.

**4.** Test the model and refine it as needed.

**5.** Prepare for the ongoing application of the model as prescribed by management.

**6.** Implement.

Operations Research

* May involve current operations or proposed developments due to expected market shifts
* May become apparent through consumer complaints or through employee suggestions
* May be a conscious effort to improve efficiency or respond to an unexpected crisis

## Linear Programming

**Linear programming:** The general problem of *optimizing* a linear function of several variables subject to a number of *constraints* that are linear in these variables and a subset of which restrict the variables to be non-negative.

NOTE: The general mathematical formulation of the *linear programming* problem is the set of matrix relationships as follows:

**Optimizing** means obtaining the best possible mathematical solution to a given set of equations.

### Math Problem

## Graphical Solution

**Further Reference in Intro to OR:**

1. Linear Programming Math - Linear Programming from OPERATIONS RESEARCH by R. PANNEERSELVAM 2nd edition (selected pages).pdf
2. Graphical Solution - Graphical Soln Introduction to OR - deterministic model JURAJ STACHO.pdf

# Simplex & Dual



## Simplex Method

### Tabular & Big M Solution

Big M

### Unbounded & Infeasible Solution

## Revised Simplex

## Dual Problem

**Further Reference in Simplex & Dual:**

Simplex Method, Big M, Infeasible & Unbounded Solution - Simplex from Introduction to Operations Research Lieberman.pdf

Duality - Duality from Introduction to OR - deterministic model Juraj Stacho.pdf

Revised Simplex - https://youtu.be/e2lHyMl1IYY

## Python Projects on Simplex, Revised Simplex Optimization

Python Code of Optimization Project 1: Revised Simplex

|  |
| --- |
| 1. """ 2. Related YouTube Video: https://youtu.be/e2lHyMl1IYY 3. """ 4. import numpy as np 5. import scipy as sp 7. c = [-3, -5] 8. A = [[1, 0], [0, 2], [3, 2]] 9. b = [4, 12, 18] 10. x0\_bounds = (0, None) 11. x1\_bounds = (0, None) 13. from scipy.optimize import linprog 14. # Solve the problem by Simplex method in Optimization 15. res = linprog(c, A\_ub=A, b\_ub=b,  bounds=(x0\_bounds, x1\_bounds), method='simplex', options={"disp": True}) 16. print(res) |

Python Code of Optimization Project : Shelf Space Optimization

|  |
| --- |
| 1. #import all relevant libraries 3. **import** pandas as pd 5. **import** numpy as np 7. **import** math 9. **from** math **import** isnan 11. **from** pulp **import** \* 13. **from** collections **import** Counter 15. #from more\_itertools import unique\_everseen   19. sales=pd.read\_csv("sales\_lift.csv",header=None) #input file 21. lift=sales.iloc[2:,1:] 23. lift=np.array(lift) 25. lift = lift.astype(np.int) # read the lifts from csv 27. brands=sales.iloc[0:1,:] 29. brands=np.array(brands) 31. brands=np.delete(brands,0) 33. brands=brands.tolist()  # read the brands from csv 35. ff=Counter(brands) 37. all\_brands=ff.items() 39. # the racks and the shelfs available 41. rack\_shelf=[[1,1,2,3],[2,4,5,6],[3,7,8,9,10]]   45. #define the optimization function 47. prob=LpProblem("SO",LpMaximize)   51. #define decision variables 53. dec\_var=LpVariable.matrix("dec\_var",(range(len(lift)),range(len(lift[0]))),0,1,LpBinary)   57. #Compute the sum product of decision variables and lifts 59. prodt\_matrix=[dec\_var[i][j]\*lift[i][j] **for** i **in** range(len(lift)) 61. **for** j **in** range(len(lift[0]))]   65. #total lift which has to be maximized sum(prodt\_matrix)   69. #define the objective function 71. prob+=lpSum(prodt\_matrix)   75. order=list(unique\_everseen(brands)) 77. order\_map = {} 79. **for** pos, item **in** enumerate(order): 81. order\_map[item] = pos 83. #brands in order as in input file 85. brands\_lift=sorted(all\_brands, key=**lambda** x: order\_map[x[0]])   89. #DEFINE CONSTRAINTS 91. #1) Each shelf can have only one product i.e. sum (each row)<=1 93. **for** i **in** range(len(lift)): 95. prob+=lpSum(dec\_var[i])<=1   99. # 2) Each product can be displayed only on a limited number of shelves i.e. Column constraints 101. #Constraints are given as 103. col\_con=[1,0,0,2,2,3,1,1] 105. dec\_var=np.array(dec\_var) 107. col\_data=[] 109. **for** j **in** range(len(brands)): 111. col\_data.append(list(zip(\*dec\_var)[j])) 113. prob+=lpSum(col\_data[j])<=col\_con[j] 115. #write the problem 117. prob.writeLP("SO.lp") 119. #solve the problem 121. prob.solve() 123. **print**("The maximum Total lift obtained is:",value(prob.objective)) # print the output 125. #print the decision variable output matrix 127. Matrix=[[0 **for** X **in** range(len(lift[0]))] **for** y **in** range(len(lift))] 129. **for** v **in** prob.variables(): 131. Matrix[int(v.name.split("\_")[2])][int(v.name.split("\_")[3])]=v.varValue 133. matrix=np.int\_(Matrix) 135. **print** ("The decision variable matrix is:") 137. **print**(matrix) |

# Transportation & Assignment



## NorthWest Corner Method

## Assignment Problem

**Further Reference in Transportation & Assignment:**

NorthWest Corner Method - NorthWest Corner Method from Introduction to Operations Research by Lieberman.pdf

Assignment problem - Assignment problem from OR topcu.pdf

## Python Projects on Optimization in Transportation & Assignment

Python Project on Optimization 1: Transportation Network for distributing products from source to destination[[3]](#footnote-3)

**Problem Description**

A company has two warehouses from which it distributes products to five carefully chosen distribution centers. The company would like to have an interactive computer program which they can run week by week to tell them which warehouse should supply which distribution center so as to minimize the costs of the whole operation. For example, suppose that at the start of a given week the company has 2050 cases at warehouse A, and 8010 cases at warehouse B, and that the distribution centers or customer points require 1000, 1800, 4000, 500, and 1350 cases respectively. Which warehouse should supply which customer point?

**Formulation**

For transportation problems, using a graphical representation of the problem is often helpful during formulation. Here is a graphical representation of The Product Distribution Problem.

Python Code of Optimization Project 3: Product Distribution Problem for the PuLP Modeller

|  |
| --- |
| # -\*- coding: utf-8 -\*-  """  The Product Distribution Problem for the PuLP Modeller  Original Authors: Antony Phillips, Dr Stuart Mitchell 2007  Adopted by: Tanmoy Das, 2018  Source code: https://github.com/openstack/deb-python-pulp/edit/master/examples/BeerDistributionProblem\_resolve.py  https://www.coin-or.org/PuLP/CaseStudies/a\_transportation\_problem.html  https://github.com/tanmoyie/Operations-Research/tree/master/Transportation  """  # Import PuLP modeler functions  from pulp import \*  # Creates a list of all the supply nodes  Warehouses = ["A", "B"]  # Creates a dictionary for the number of units of supply for each supply node  supply = {"A": 2050,  "B": 8010}  # Creates a list of all demand nodes  CustomerPoint = ["1", "2", "3", "4", "5"]  # Creates a dictionary for the number of units of demand for each demand node  demand = {"1":1000,  "2":1800,  "3":4000,  "4":500,  "5":1350,}  # Creates a list of costs of each transportation path  costs = [ #CustomerPoint  #1 2 3 4 5  [2,4,5,2,1],#A Warehouses  [3,1,3,2,3] #B  ]  # The cost data is made into a dictionary  costs = makeDict([Warehouses,CustomerPoint],costs,0)  # Creates the 'prob' variable to contain the problem data  prob = LpProblem("Product Distribution Problem",LpMinimize)  # Creates a list of tuples containing all the possible routes for transport  Routes = [(w,b) for w in Warehouses for b in CustomerPoint]  # A dictionary called 'Vars' is created to contain the referenced variables(the routes)  vars = LpVariable.dicts("Route",(Warehouses,CustomerPoint),0,None,LpInteger)  # The objective function is added to 'prob' first  prob += lpSum([vars[w][b]\*costs[w][b] for (w,b) in Routes]), "Sum\_of\_Transporting\_Costs"  # The supply maximum constraints are added to prob for each supply node (warehouse)  for w in Warehouses:  prob += lpSum([vars[w][b] for b in CustomerPoint])<=supply[w], "Sum\_of\_Products\_out\_of\_Warehouse\_%s"%w  # The demand minimum constraints are added to prob for each demand node (customer)  # These constraints are stored for resolve later  customer\_demand\_constraint = {}  for b in CustomerPoint:  constraint = lpSum([vars[w][b] for w in Warehouses])>=demand[b]  prob += constraint, "Sum\_of\_Products\_into\_customer\_%s"%b  customer\_demand\_constraint[b] = constraint  # The problem data is written to an .lp file  prob.writeLP("ProductDistributionProblem.lp")  for demand in range(500, 601, 10):  # reoptimise the problem by increasing demand at customer '1'  # note the constant is stored as the LHS constant not the RHS of the constraint  customer\_demand\_constraint['1'].constant = - demand  # The problem is solved using PuLP's choice of Solver  prob.solve()  # The status of the solution is printed to the screen  print("Status:", LpStatus[prob.status])  # Each of the variables is printed with it's resolved optimum value  for v in prob.variables():  print(v.name, "=", v.varValue)  # The optimised objective function value is printed to the screen  print("Total Cost of Transportation = ", value(prob.objective)) |

Table 1: Output

|  |
| --- |
| runfile('G:/Github Tanmoy Das/Operations-Research/Transportation/transportation\_problem\_PuLP\_example\_of\_product\_distribution\_from\_warehouse\_to\_customer.py', wdir='G:/Github Tanmoy Das/Operations-Research/Transportation')  Status: Optimal  Route\_A\_1 = 500.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 200.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 300.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17150.0  Status: Optimal  Route\_A\_1 = 510.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 190.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 310.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17170.0  Status: Optimal  Route\_A\_1 = 520.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 180.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 320.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17190.0  Status: Optimal  Route\_A\_1 = 530.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 170.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 330.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17210.0  Status: Optimal  Route\_A\_1 = 540.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 160.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 340.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17230.0  Status: Optimal  Route\_A\_1 = 550.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 150.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 350.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17250.0  Status: Optimal  Route\_A\_1 = 560.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 140.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 360.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17270.0  Status: Optimal  Route\_A\_1 = 570.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 130.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 370.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17290.0  Status: Optimal  Route\_A\_1 = 580.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 120.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 380.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17310.0  Status: Optimal  Route\_A\_1 = 590.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 110.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 390.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17330.0  Status: Optimal  Route\_A\_1 = 600.0  Route\_A\_2 = 0.0  Route\_A\_3 = 0.0  Route\_A\_4 = 100.0  Route\_A\_5 = 1350.0  Route\_B\_1 = 0.0  Route\_B\_2 = 1800.0  Route\_B\_3 = 4000.0  Route\_B\_4 = 400.0  Route\_B\_5 = 0.0  Total Cost of Transportation = 17350.0 |

Python Code of Optimization Project 4: Travelling Salesman Problem

|  |
| --- |
| 1. """ 2. Running Code: https://www.kaggle.com/tanmoyie/traveling-salesman-problem 3. Source code: https://developers.google.com/optimization/routing/tsp 4. Google Map: https://drive.google.com/open?id=18k6uA3KdLJGc2qQGUEXbtfG93KJj8Lty&usp=sharing 5. Detail video: <https://youtu.be/e2lHyMl1IYY> 6. """ 7. **from** ortools.constraint\_solver **import** pywrapcp 8. **from** ortools.constraint\_solver **import** routing\_enums\_pb2 10. # Distance callback 11. **class** CreateDistanceCallback(object): 12. """Create callback to calculate distances between points.""" 13. **def** \_\_init\_\_(self): 14. """Array of distances between points.""" 16. self.matrix = [ 17. [  0, 290, 250,  230,  190,  334, 365,   40], # Dhaka 18. [290,   0, 337,  453,  396,  560, 581,  244], # Syhlet 19. [250, 337,   0,  495,  396,  540, 120,  240], # Chittagonj 20. [230, 453, 495,    0,  360,  150, 595,  242], # Rajshahi 21. [190, 396, 396,  360,    0,  356, 496,  253], # Jossore 22. [334, 560, 540,  150,  356,    0, 674,  275], # Dinajpur 23. [365, 581, 120,  595,  496,  674,   0,  397], # Coxsbazar 24. [40,  244, 240,  242,  253,  275, 397,    0]] # Narsingdi 25. # distance between Dhaka to Syhlet is 290kms and so on 26. **def** Distance(self, from\_node, to\_node): 27. **return** int(self.matrix[from\_node][to\_node]) 28. **def** main(): 29. # The order of the cities in the array is the following: 30. # Cities 31. city\_names = ["Dhaka", "Syhlet", "Chittagonj", "Rajshahi", "Jossore", "Dinajpur", "Coxsbazar", 32. "Narsingdi"] 33. tsp\_size = len(city\_names) 34. num\_routes = 1    # The number of routes, which is 1 in the TSP. 35. # Nodes are indexed from 0 to tsp\_size - 1. The depot is the starting node of the route. 36. depot = 0 38. # Create routing model 39. **if** tsp\_size > 0: 40. routing = pywrapcp.RoutingModel(tsp\_size, num\_routes, depot) 41. search\_parameters = pywrapcp.RoutingModel.DefaultSearchParameters() 43. # Create the distance callback, which takes two arguments (the from and to node indices) 44. # and returns the distance between these nodes. 45. dist\_between\_nodes = CreateDistanceCallback() 46. dist\_callback = dist\_between\_nodes.Distance 47. routing.SetArcCostEvaluatorOfAllVehicles(dist\_callback) 48. # Solve, returns a solution if any. 49. assignment = routing.SolveWithParameters(search\_parameters) 50. **if** assignment: 51. # Solution cost. 52. **print** ("Total distance: " + str(assignment.ObjectiveValue()) + " miles\n") 53. # Inspect solution. 54. # Only one route here; otherwise iterate from 0 to routing.vehicles() - 1 55. route\_number = 0 56. index = routing.Start(route\_number) # Index of the variable for the starting node. 57. route = '' 58. **while** **not** routing.IsEnd(index): 59. # Convert variable indices to node indices in the displayed route. 60. route += str(city\_names[routing.IndexToNode(index)]) + ' -> ' 61. index = assignment.Value(routing.NextVar(index)) 62. route += str(city\_names[routing.IndexToNode(index)]) 63. **print** ("Route:\n\n" + route) 64. **else**: 65. **print** ('No solution found.') 66. **else**: 67. **print** ('Specify an instance greater than 0.') 69. **if** \_\_name\_\_ == '\_\_main\_\_': 70. main() |

# Optimization in Machine Learning/ Data Science



## Linear Regression

## Robust Regression

## Support Vector Machine

**Further Reference in Machine Learning & Optimization:**

Linear Regression - Follow Class Lecture

Robust Regression - Follow Class Lecture

Support Vector Machine - Follow Class Lecture

## Python Projects on Machine Learning Optimization

Python Code of Optimization Project : Linear Regression (plot\_ols.py)

|  |
| --- |
| 1. #!/usr/bin/python 2. # -\*- coding: utf-8 -\*- 4. """ 5. ========================================================= 6. Linear Regression Example 7. ========================================================= 8. This example uses the only the first feature of the `diabetes` dataset, in 9. order to illustrate a two-dimensional plot of this regression technique. The 10. straight line can be seen in the plot, showing how linear regression attempts 11. to draw a straight line that will best minimize the residual sum of squares 12. between the observed responses in the dataset, and the responses predicted by 13. the linear approximation. 15. The coefficients, the residual sum of squares and the variance score are also 16. calculated. 18. """ 19. **print**(\_\_doc\_\_) 20. # Code source: Jaques Grobler 22. **import** matplotlib.pyplot as plt 23. **import** numpy as np 24. **from** sklearn **import** datasets, linear\_model 25. **from** sklearn.metrics **import** mean\_squared\_error, r2\_score 27. # Load the diabetes dataset 28. diabetes = datasets.load\_diabetes() 29. # Use only one feature 30. diabetes\_X = diabetes.data[:, np.newaxis, 2] 31. # Split the data into training/testing sets 32. diabetes\_X\_train = diabetes\_X[:-20] 33. diabetes\_X\_test = diabetes\_X[-20:] 34. # Split the targets into training/testing sets 35. diabetes\_y\_train = diabetes.target[:-20] 36. diabetes\_y\_test = diabetes.target[-20:] 38. # Create linear regression object 39. regr = linear\_model.LinearRegression() 40. # Train the model using the training sets 41. regr.fit(diabetes\_X\_train, diabetes\_y\_train) 42. # Make predictions using the testing set 43. diabetes\_y\_pred = regr.predict(diabetes\_X\_test) 45. # The coefficients 46. **print**('Coefficients: \n', regr.coef\_) 47. # The mean squared error 48. **print**("Mean squared error: %.2f" 49. % mean\_squared\_error(diabetes\_y\_test, diabetes\_y\_pred)) 50. # Explained variance score: 1 is perfect prediction 51. **print**('Variance score: %.2f' % r2\_score(diabetes\_y\_test, diabetes\_y\_pred)) 53. # Plot outputs 54. plt.scatter(diabetes\_X\_test, diabetes\_y\_test,  color='black') 55. plt.plot(diabetes\_X\_test, diabetes\_y\_pred, color='blue', linewidth=3) 57. plt.xticks(()) 58. plt.yticks(()) 59. plt.show() |

Python Code of Optimization Project : Robust Regression compared to Linear Regression (plot\_ransac.py)

|  |
| --- |
| 1. """ 2. Robust linear model estimation using RANSAC 3. In this example we see how to robustly fit a linear model to faulty data using 4. the RANSAC algorithm. 5. Source Code: http://scikit-learn.org/stable/auto\_examples/linear\_model/plot\_ransac.html 6. """ 7. **import** numpy as np 8. **from** matplotlib **import** pyplot as plt 9. **from** sklearn **import** linear\_model, datasets 11. n\_samples = 1000 12. n\_outliers = 50 13. X, y, coef = datasets.make\_regression(n\_samples=n\_samples, n\_features=1, 14. n\_informative=1, noise=10, 15. coef=True, random\_state=0) 17. # Add outlier data 18. np.random.seed(0) 19. X[:n\_outliers] = 3 + 0.5 \* np.random.normal(size=(n\_outliers, 1)) 20. y[:n\_outliers] = -3 + 10 \* np.random.normal(size=n\_outliers) 22. # Fit line using all data 23. lr = linear\_model.LinearRegression() 24. lr.fit(X, y) 26. # Robustly fit linear model with RANSAC algorithm 27. ransac = linear\_model.RANSACRegressor() 28. ransac.fit(X, y) 29. inlier\_mask = ransac.inlier\_mask\_ 30. outlier\_mask = np.logical\_not(inlier\_mask) 32. # Predict data of estimated models 33. line\_X = np.arange(X.min(), X.max())[:, np.newaxis] 34. line\_y = lr.predict(line\_X) 35. line\_y\_ransac = ransac.predict(line\_X) 37. # Compare estimated coefficients 38. **print**("Estimated coefficients (true, linear regression, RANSAC):") 39. **print**(coef, lr.coef\_, ransac.estimator\_.coef\_) 40. lw = 2 41. plt.scatter(X[inlier\_mask], y[inlier\_mask], color='yellowgreen', marker='.', 42. label='Inliers') 43. plt.scatter(X[outlier\_mask], y[outlier\_mask], color='black', marker='+', 44. label='Outliers') 45. plt.plot(line\_X, line\_y, color='navy', linewidth=lw, label='Linear regressor') 46. plt.plot(line\_X, line\_y\_ransac, color='black', linestyle='--',, linewidth=lw, 47. label='RANSAC regressor') 48. plt.legend(loc='lower right') 49. plt.xlabel("Input") 50. plt.ylabel("Response") 51. plt.show() |

# Network Optimization



## Min Cost

## Max Flow

## Minimum Spanning Tree

**Further Reference in Network Optimization:**

Min Cost, Max Flow - Network Optimization from Introduction to OR - deterministic model Juraj Stacho.pdf

Shortest Path, MST - Network Optimization from Introduction to Operations Research by Lieberman.pdf

## Python Projects on Network Optimization

Python Code of Optimization Project 7: Min Cost Flow problem

|  |
| --- |
| 1. """ 2. Running kernel: https://www.kaggle.com/tanmoyie/min-cost-flow-google-developer 3. Source: https://developers.google.com/optimization/flow/mincostflow 5. """ 7. # """From Bradley, Hax, and Magnanti, 'Applied Mathematical Programming', figure 8.1.""" 9. **from** \_\_future\_\_ **import** print\_function 10. **from** ortools.graph **import** pywrapgraph 12. **def** main(): 13. """MinCostFlow simple interface example.""" 15. # Define four parallel arrays: start\_nodes, end\_nodes, capacities, and unit costs 16. # between each pair. For instance, the arc from node 0 to node 1 has a 17. # capacity of 15 and a unit cost of 4. 19. start\_nodes = [0, 0, 1, 1, 1, 2, 2, 3, 4] 20. end\_nodes = [1, 2, 2, 3, 4, 3, 4, 4, 2] 21. capacities = [15, 8, 20, 4, 10, 15, 5, 20, 4] 22. unit\_costs = [4, 4, 2, 2, 6, 1, 3, 2, 3] 24. # Define an array of supplies at each node. 26. supplies = [20, 0, 0, -5, -15] 28. # Instantiate a SimpleMinCostFlow solver. 29. min\_cost\_flow = pywrapgraph.SimpleMinCostFlow() 31. # Add each arc. 32. **for** i **in** range(0, len(start\_nodes)): 33. min\_cost\_flow.AddArcWithCapacityAndUnitCost(start\_nodes[i], end\_nodes[i], 34. capacities[i], unit\_costs[i]) 36. # Add node supplies. 38. **for** i **in** range(0, len(supplies)): 39. min\_cost\_flow.SetNodeSupply(i, supplies[i])  42. # Find the minimum cost flow between node 0 and node 4. 43. **if** min\_cost\_flow.Solve() == min\_cost\_flow.OPTIMAL: 44. **print**('Minimum cost:', min\_cost\_flow.OptimalCost()) 45. **print**('') 46. **print**('  Arc    Flow / Capacity  Cost') 47. **for** i **in** range(min\_cost\_flow.NumArcs()): 48. cost = min\_cost\_flow.Flow(i) \* min\_cost\_flow.UnitCost(i) 49. **print**('%1s -> %1s   %3s  / %3s       %3s' % ( 50. min\_cost\_flow.Tail(i), 51. min\_cost\_flow.Head(i), 52. min\_cost\_flow.Flow(i), 53. min\_cost\_flow.Capacity(i), 54. cost)) 55. **else**: 56. **print**('There was an issue with the min cost flow input.') 58. **if** \_\_name\_\_ == '\_\_main\_\_': 59. main() |

Python Code of Optimization Project : Max Flow

|  |
| --- |
| """From Taha 'Introduction to Operations Research', example 6.4-2."""  # https://developers.google.com/optimization/flow/maxflow  from \_\_future\_\_ import print\_function  from ortools.graph import pywrapgraph  def main():  """MaxFlow simple interface example."""  # Define three parallel arrays: start\_nodes, end\_nodes, and the capacities  # between each pair. For instance, the arc from node 0 to node 1 has a  # capacity of 20.  start\_nodes = [0, 0, 0, 1, 1, 2, 2, 3, 3]  end\_nodes = [1, 2, 3, 2, 4, 3, 4, 2, 4]  capacities = [20, 30, 10, 40, 30, 10, 20, 5, 20]  # Instantiate a SimpleMaxFlow solver.  max\_flow = pywrapgraph.SimpleMaxFlow()  # Add each arc.  for i in range(0, len(start\_nodes)):  max\_flow.AddArcWithCapacity(start\_nodes[i], end\_nodes[i], capacities[i])  # Find the maximum flow between node 0 and node 4.  if max\_flow.Solve(0, 4) == max\_flow.OPTIMAL:  print('Max flow:', max\_flow.OptimalFlow())  print('')  print(' Arc Flow / Capacity')  for i in range(max\_flow.NumArcs()):  print('%1s -> %1s %3s / %3s' % (  max\_flow.Tail(i),  max\_flow.Head(i),  max\_flow.Flow(i),  max\_flow.Capacity(i)))  print('Source side min-cut:', max\_flow.GetSourceSideMinCut())  print('Sink side min-cut:', max\_flow.GetSinkSideMinCut())  else:  print('There was an issue with the max flow input.')  if \_\_name\_\_ == '\_\_main\_\_':  main() |

Python Code of Optimization Project 9: Airlines Network Optimization[[4]](#footnote-4)

|  |
| --- |
| 1. # -\*- coding: utf-8 -\*- 2. """ 3. Created on Tue Jun  5 20:09:14 2018 5. @adopted by: Tanmoy Das 6. Earlier version: https://www.analyticsvidhya.com/blog/2018/04/introduction-to-graph-theory-network-analysis-python-codes/ 8. """ 10. # import the libraries 11. **import** numpy as np # linear algebra 12. **import** pandas as pd # data processing, CSV file I/O (e.g. pd.read\_csv) 13. **import** os 15. # load the dataset 16. data = pd.read\_csv('airlines\_network\_optimization.csv') # download the csv file in your local directory and play with it. 18. # data.shape 19. # converting sched\_dep\_time to 'std' - Scheduled time of departure 20. data['std'] = data.sched\_dep\_time.astype(str).str.replace('(\d{2}$)', '') + ':' + data.sched\_dep\_time.astype(str).str.extract('(\d{2}$)', expand=False) + ':00' 21. # converting sched\_arr\_time to 'sta' - Scheduled time of arrival 22. data['sta'] = data.sched\_arr\_time.astype(str).str.replace('(\d{2}$)', '') + ':' + data.sched\_arr\_time.astype(str).str.extract('(\d{2}$)', expand=False) + ':00' 23. # converting dep\_time to 'atd' - Actual time of departure 24. data['atd'] = data.dep\_time.fillna(0).astype(np.int64).astype(str).str.replace('(\d{2}$)', '') + ':' + data.dep\_time.fillna(0).astype(np.int64).astype(str).str.extract('(\d{2}$)', expand=False) + ':00' 25. # converting arr\_time to 'ata' - Actual time of arrival 26. data['ata'] = data.arr\_time.fillna(0).astype(np.int64).astype(str).str.replace('(\d{2}$)', '') + ':' + data.arr\_time.fillna(0).astype(np.int64).astype(str).str.extract('(\d{2}$)', expand=False) + ':00' 27. data['date'] = pd.to\_datetime(data[['year', 'month', 'day']]) 28. # finally we drop the columns we don't need 29. data = data.drop(columns = ['year', 'month', 'day']) 31. **import** networkx as nx 32. FG = nx.from\_pandas\_edgelist(data, source='origin', target='dest', edge\_attr=True,) 33. # detail documentation of networkx https://networkx.github.io/documentation/networkx-1.7/reference/generated/networkx.drawing.nx\_pylab.draw\_networkx.html 34. FG.nodes() 35. FG.edges() 36. nx.draw\_networkx(FG, with\_labels=True,node\_size=600, node\_color='y') # Quick view of the Graph. As expected we see 3 very busy airports 38. nx.algorithms.degree\_centrality(FG) # Notice the 3 airports from which all of our 100 rows of data originates 39. nx.density(FG) # Average edge density of the Graphs 40. nx.average\_shortest\_path\_length(FG) # Average shortest path length for ALL paths in the Graph 41. nx.average\_degree\_connectivity(FG) # For a node of degree k - What is the average of its neighbours' degree? 43. # Let us find all the paths available 44. **for** path **in** nx.all\_simple\_paths(FG, source='JAX', target='DFW'): 45. **print**(path) 46. # Let us find the dijkstra path from JAX to DFW. 47. # You can read more in-depth on how dijkstra works from this resource - https://courses.csail.mit.edu/6.006/fall11/lectures/lecture16.pdf 48. dijpath = nx.dijkstra\_path(FG, source='JAX', target='DFW') 49. dijpath 50. # Let us try to find the dijkstra path weighted by airtime (approximate case) 51. shortpath = nx.dijkstra\_path(FG, source='JAX', target='DFW', weight='air\_time') 52. shortpath |

Airlines\_network\_optimization.csv (first 10 records)

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  | | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | | year | month | day | dep\_time | sched\_dep\_time | dep\_delay | arr\_time | sched\_arr\_time | arr\_delay | carrier | flight | tailnum | origin | dest | air\_time | distance | | 2013 | 2 | 26 | 1807 | 1630 | 97 | 1956 | 1837 | 79 | EV | 4411 | N13566 | EWR | MEM | 144 | 946 | | 2013 | 8 | 17 | 1459 | 1445 | 14 | 1801 | 1747 | 14 | B6 | 1171 | N661JB | LGA | FLL | 147 | 1076 | | 2013 | 2 | 13 | 1812 | 1815 | -3 | 2055 | 2125 | -30 | AS | 7 | N403AS | EWR | SEA | 315 | 2402 | | 2013 | 4 | 11 | 2122 | 2115 | 7 | 2339 | 2353 | -14 | B6 | 97 | N656JB | JFK | DEN | 221 | 1626 | | 2013 | 8 | 5 | 1832 | 1835 | -3 | 2145 | 2155 | -10 | AA | 269 | N3EYAA | JFK | SEA | 358 | 2422 | | 2013 | 6 | 30 | 1500 | 1505 | -5 | 1751 | 1650 | 61 | UA | 685 | N424UA | LGA | ORD | 116 | 733 | | 2013 | 2 | 14 | 1442 | 1445 | -3 | 1833 | 1747 | 46 | UA | 346 | N446UA | EWR | MIA | 200 | 1085 | | 2013 | 7 | 25 | 752 | 755 | -3 | 1037 | 1057 | -20 | DL | 2395 | N909DL | LGA | PBI | 140 | 1035 | | 2013 | 7 | 10 | 557 | 600 | -3 | 725 | 715 | 10 | MQ | 3267 | N542MQ | EWR | ORD | 113 | 719 | |

Table 2: Output of Airlines Network Optimization

|  |
| --- |
||

# Integer Programming



**Further Reference in Integer Programming:**

Integer Programming - Integer Programming from OPERATIONS RESEARCH by R. PANNEERSELVAM 2nd edition (selected pages).pdf

1. Latter two chapters is estimated to be covered by another instructor, hence, skipped for this document! [↑](#footnote-ref-1)
2. A technical writeup [↑](#footnote-ref-2)
3. Source: https://www.coin-or.org/PuLP/CaseStudies/a\_transportation\_problem.html [↑](#footnote-ref-3)
4. Formatting & Color help from http://www.planetb.ca/syntax-highlight-word [↑](#footnote-ref-4)