40.016 The Analytics Edge

Prescriptive analytics with Julia (Part 1)

Stefano Galelli

SUTD

Term 6, 2020

Outline

- Julia
- IJulia
- JuMP
- Revenue management (aviation industry)
- Discount allocation
- Pricing (optional)

Julia

Key features:

- General-purpose programming language
- Dynamic-type system
- Open source

Julia

Key features (cont'd):

- Performance approaching that of statically-typed languages (e.g., C/C++)
- Interfaces with other languages, such as R and Python
- Supports parallel and distributed computing
- Built-in package manager

Julia

History:

- Work on Julia started in 2009
- Julia Computing founded in 2015
- In 2017, Julia was used for petascale computing (i.e., 10^{15} floating point operations per second)
- Julia 1.0 released in 2018

Julia

Editors:

- Juno
- VSCode
- Jupyter Notebooks
  - The easiest way to install Jupyter and Python simultaneously is to use Anaconda
  - Requires IJulia (package that provides the connection between Jupyter and Julia)

IJulia and Jupyter

How to launch Jupyter for Julia?

Option 1: Type `jupyter notebook` in the terminal

Option 2: Type the following commands in Julia (`julia>` prompt)

```
using IJulia
notebook()
```

JuMP

Key features:

- Modeling language for mathematical optimization embedded in Julia
- Supports many solvers (both open-source and commercial)
- See the Installation Guide for a list of solvers
- Several classes of problems (LP, MILP, nonlinear programs etc.)

JuMP

Workflow for creating a model:

1. Load JuMP and a solver / optimizer (e.g., GLPK)
2. Create a model with the `Model()` function
3. Add variables (`@variable()`)
4. Add objective (`@objective()`)
5. Add constraints, if any (`@constrain()`)

JuMP

Workflow for solving a problem:

1. Use the `optimize!()` function
2. Check the status of the solution (`termination_status()`)

JuMP

Example 1:

$$\begin{aligned} \min_{x,y} \quad & x + y \\ \text{s.t.} \quad & x + y \leq 1 \\ & x, y \geq 0 \\ & x, y \in \mathbb{R} \end{aligned}$$

File `Example 1.ipynb`.

Example 2:

$$
\begin{aligned}
\min_{x,y} \quad & c^T x \\
\text{s.t.} \quad & Ax = b \\
& x \geq 0 \\
& x \in \mathbb{R}^n
\end{aligned}
$$

File `Example 2.ipynb`.

Example 3:

$$
\begin{aligned}
\min_{x,y} \quad & c^T x + d^T y \\
\text{s.t.} \quad & Ax + By = f \\
& x, y \geq 0 \\
& x \in \mathbb{R}^n, y \in \mathbb{Z}^p
\end{aligned}
$$

File `Example 3.ipynb`.

Revenue management (aviation industry)

- Deregulation in the aviation industry (late 70s)
- Opportunity for developing and implementing revenue management techniques
- Key challenge: *sell the right product to the right customer at the right time for the very right price.*

Revenue management (aviation industry)

Key features of the aviation industry:

- Fixed capacity (i.e., number of seats)
- High fixed costs but low variable costs
- Variety of customer types
- Ability to sell tickets to customers without seeing what other customers paid

Revenue management (aviation industry)

Common aspects:

- Traffic management
- Discount allocation
- Pricing
- Overbooking

Discount allocation

Network

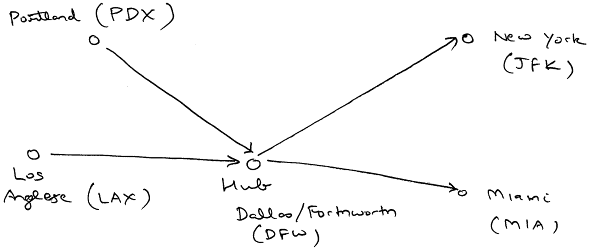

Figure 1: Network centred in Dallas/Fort Worth International Airport (DFW).

Discount allocation

Idea: We offer two products

1. More expensive tickets that can be purchased anytime, with no restrictions
2. Cheaper tickets, to be bought at least 3 weeks in advance, with penalties for changes

Discount allocation

Assumptions:

1. Prices are set by the marketing department
2. We are interested in deciding **how many seats** to reserve in each fare class and itinerary
3. We consider three flights: LAX $\rightarrow$ DFW, LAX $\rightarrow$ JFK, and DFW $\rightarrow$ JFK

Discount allocation

Data

Table 1: Capacity of LAX $\rightarrow$ DFW and DFW $\rightarrow$ JFK flights.

Flight	Capacity
LAX $\rightarrow$ DFW	300
DFW $\rightarrow$ JFK	200

Discount allocation

Data

Table 2: Revenue and demand for each fare of LAX $\rightarrow$ DFW, LAX $\rightarrow$ JFK, and DFW $\rightarrow$ JFK flights. The last column reports the decision variables for each leg.

Flight	Fare	Revenue (USD)	Demand	Decision variable
LAX $\rightarrow$ DFW	Regular	100	20	x_1
	Discount	90	40	x_2
	Saver	80	60	x_3
LAX $\rightarrow$ JFK	Regular	215	80	y_1
	Discount	185	60	y_2
	Saver	145	70	y_3
DFW $\rightarrow$ JFK	Regular	140	20	z_1
	Discount	120	20	z_2
	Saver	100	30	z_3

Discount allocation

Problem formulation

$$
\begin{aligned}
\max \quad & 100x_1 + 90x_2 + 80x_3 + 215y_1 + 185y_2 + 145y_3 + 140z_1 + 120z_2 + 100z_3 \\
\text{s.t.} \quad & x_1 + x_2 + x_3 + y_1 + y_2 + y_3 \leq 300 \\
& y_1 + y_2 + y_3 + z_1 + z_2 + z_3 \leq 200 \\
& 0 \leq x_1 \leq 20 \\
& 0 \leq x_2 \leq 40 \\
& 0 \leq x_3 \leq 60 \\
& 0 \leq y_1 \leq 80 \\
& 0 \leq y_2 \leq 60 \\
& 0 \leq y_3 \leq 70 \\
& 0 \leq z_1 \leq 20 \\
& 0 \leq z_2 \leq 20 \\
& 0 \leq z_3 \leq 30
\end{aligned}
$$

File `seats.ipynb`.

Discount allocation

The optimal solution is: $x_1^* = 20$, $x_2^* = 40$, $x_3^* = 60$, $y_1^* = 80$, $y_2^* = 60$, $y_3^* = 40$, $z_1^* = 20$, $z_2^* = 0$, $z_3^* = 0$, with a total revenue of USD 47,300.

A **greedy method** would start by allocating seats to the customers yielding the highest revenue, and then focus on the remaining customers. Such method would yield a revenue of USD 45,800.

Pricing (optional)

Pricing products with a Multinomial Logit model

Consider a set of products $\{1, \dots, n\}$ and an outside option denoted by 0. The goal of the decision-maker is to price products $1, \dots, n$ through the following optimization problem:

$$\max_{p_1 \geq 0, \dots, p_n \geq 0} \sum_{i=1}^{n} \Big(\underbrace{p_i}_{\text{price}} - \underbrace{c_i}_{\text{cost}} \Big) \underbrace{\mathbb{P}(\text{choosing product } i \text{ given prices } \bar{P})}_{\substack{\text{Obtained from} \\ \text{logit model}}},$$

where

$$\mathbb{P}(\text{choosing product } i \text{ given prices } \bar{P}) = \frac{e^{v_i - p_i}}{\sum_{j=0}^{n} e^{v_j - p_j}},$$

where p_i is the price of the i-th product and v_i its corresponding evaluation.

Pricing (optional)

Pricing products with a Multinomial Logit model

The optimization problem can be rewritten as:

$$\max_{p_1 \geq 0, \ldots, p_n \geq 0} \sum_{i=1}^{n} \frac{(p_i - c_i) e^{v_i - p_i}}{\sum_{j=0}^{n} e^{v_j - p_j}},$$

where $v_0 = p_0 = 0$ are the evaluation and price of the default (outside) option.

This is not an easy problem to solve. The problem can be re-formulated and solved using choice probability variables (see the derivation in the teaching notes). The implementation is provided in the file `pricing.ipynb`.

References

Julia (with links)

- *Julia*
- *IJ*
- *IJulia*
- *JuMP*

References

Revenue management

- Teaching notes.
- Bertsimas, D., Allison, K. O., & Pulleyblank, W. R. (2016). *The Analytics Edge*. Dynamic Ideas LLC.
- Smith, B. C., Leimkuhler, J. F., & Darrow, R. M. (1992). Yield management at American airlines. *Interfaces*, 22(1), 8-31.