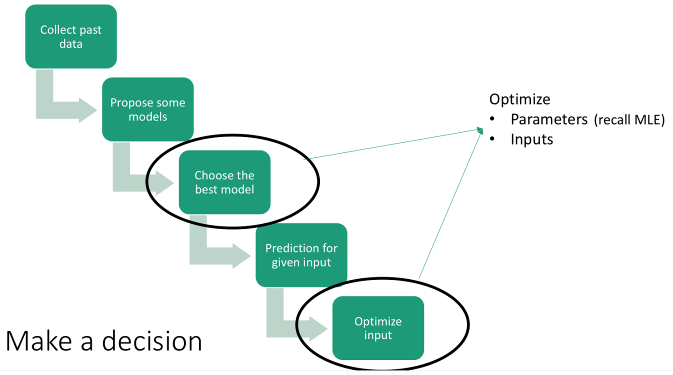
Lecture 15

§1 Optimization Basics Introduction

1. 应用情景



2. 为什么 study optimization

1° Optimization 是 Data Science 的基础

- Modeling
- Solution method

2° Optimization 可以帮助

- 深入理解 probability / statistics / ML approach
- 解释算法
- 开发新算法

Wikipedia:

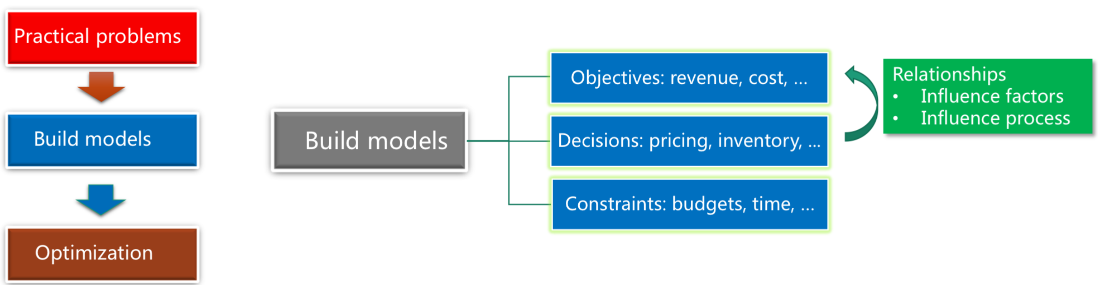
"An optimization problem consists of maximizing or minimizing a real function by systematically choosing input values from within an allowed set and computing the value of the function."

Bernoulli RV with parameter p.
p shall be within [0,1].

To recommend a product, you need to recommend products that you have.

§2 Concepts: how to formulate the problem.

1. 基本步骤



X: decisions (input)

R: objective (output)

$R = f(x)$

2. Problem formulation (问题公式化)

1° An optimization problem is specified by

Objective function

$$\begin{aligned} \text{minimize} \quad & \boxed{f(x)} \\ \text{subject to} \quad & \boxed{\begin{aligned} & g_i(x) \le 0, i = 1, \ldots, m \\ & h_i(x) = 0, i = 1, \ldots, n \end{aligned}} \end{aligned}$$

Constraints

2° An optimization problem is to:

① minimize or maximize an objective

② subject to (服从) constraints (等式/不等式(weak inequality: ≤/≥)

例: **Problem**

- Bob is in a pie-eating contest that lasts **1 hour.** Each torte that he eats takes 2 minutes. Each apple pie that he eats takes 3 minutes. He receives 4 points for each torte and 5 points for each pie. How should Bob eat so as to get the most points?

use 2 minutes
gain 4 points

use 3 minutes
gain 5 points

Decisions

Let x be the number of tortes eaten by Bob.
Let y be the number of pies eaten by Bob.

Objective

Maximize

z = 4x + 5y (objective function)

Constraints

subject to

2x + 3y ≤ 60 (time constraint)
x ≥ 0 ; y ≥ 0 (non-negativity constraints)

Problem

The optimization problem is formulated as
Maximize z = 4x + 5y
Subject to
2x + 3y ≤ 60
x ≥ 0 ; y ≥ 0
where x be the number of tortes eaten and y be the number of pies eaten.

A **feasible solution** satisfies all the constraints.
x = 10, y = 10 is feasible; x = 10, y = 15 is infeasible.
An **optimal solution** is the best feasible solution that optimizes the objective.
The optimal solution is x = 30, y = 0, z = 120.

3. terminologies

1° Decision variables (决策变量): e.g. x and y

可以控制并用于提升 objective 的量

2° Constraints (约束条件): e.g. $2x+3y \le 60$, $x \ge 0$, $y \ge 0$
对于 decision variables 取值的限制

3° Objective function (目标函数): e.g. $4x+5y$

① 值被用于 rank alternatives

② 要将目标函数最大化/最小化.

③ 例: maximize reward, minimize cost

案例分析 1: MLE

Problem

- Let's revisit an MLE example.
- Suppose we have a random sample

$$X_1, X_2, \ldots, X_n$$

where $X_i = 0$ if a randomly selected student does not own a sports car, and $X_i = 1$ if a randomly selected student does own a sports car.

Assuming that they are independent Bernoulli random variables with unknown parameter, find the maximum likelihood estimator of the proportion of students who own a sports car.

In this optimization problem

- What is the objective function?

  max log-likelihood function (*p*)

- What is the decision variable?

  p

- What is the constraint?

$$0 \le p \le 1$$

Answer

If the X_i are independent Bernoulli random variables with unknown parameter p, then the probability mass function of each X_i is:

$$f(x_i;p) = p^{x_i}(1-p)^{1-x_i}$$

for $x_i = 0$ or 1 and $0 < p < 1$. Therefore, the likelihood function $L(p)$ is, by definition:

$$L(p) = \prod_{i=1}^{n} f(x_i;p) = p^{x_1}(1-p)^{1-x_1} \times p^{x_2}(1-p)^{1-x_2} \times \cdots \times p^{x_n}(1-p)^{1-x_n}$$

for $0 < p < 1$. Simplifying, by summing up the exponents, we get :

$$L(p) = p^{\sum x_i}(1-p)^{n-\sum x_i}$$

In this case, the natural logarithm of the likelihood function is:

$$\log L(p) = (\sum x_i)\log(p) + (n - \sum x_i)\log(1-p)$$

Now, taking the derivative of the log likelihood, and setting to 0, we get:

$$\sum x_i - np = 0$$

Now, all we have to do is solve for p. In doing so, you'll want to make sure that you always put a hat ("^") on the parameter, in this case p, to indicate it is an estimate:

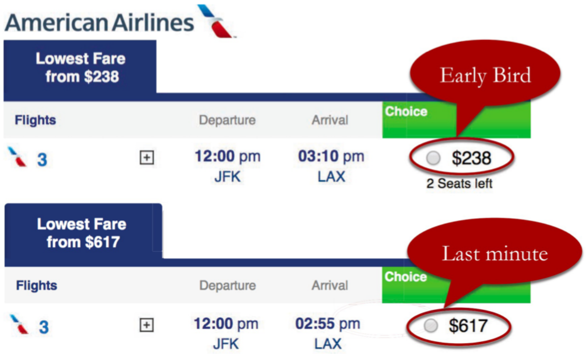
$$\hat{p} = \frac{\sum_{i=1}^{n} x_i}{n}$$

Note that, we should technically verify that we indeed did **obtain a maximum**. We can do that by *verifying that the second derivative of the log likelihood with respect to p* is negative.

It is, but you might want to do the work to convince yourself!

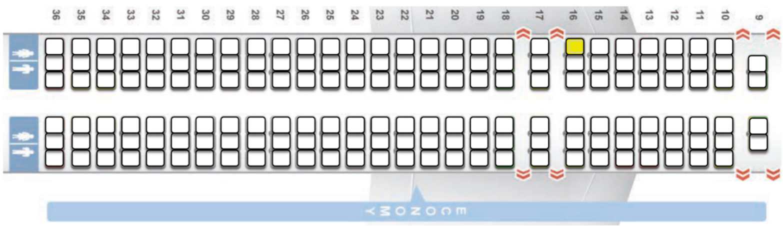
案例分析2: Revenue management

Airline companies: How to maximize the revenue?



Constraints

Seat Compacity **Constraint**: A total 166 economy seats



Decisions

- Let's **assume** that the prices for different types are already fixed.
  - This may be because that their competitors set such prices.

- **How** many discount seats to sell to maximize revenue?

		Price	Demand	Seats to Sell
JFK LAX	Regular	617	100	
	Discount	238	150	

Capacity 166

Objective

What is our **objective function**?

- Maximizing total airline revenue
- Revenue from each type of seat is equal to the number of that type of seat sold times the seat price

$$max\ \ 617\,R + \ 238\,D$$

		Price	Demand	Seats to Sell
JFK LAX	Regular	617	100	
	Discount	238	150	

Capacity 166

Decisions

Problem: Find the optimal number of **discounted seats** and r**egular seats** to sell to maximize revenue

Let's formulate the optimization problem mathematically

What are our **decision variables**?

- Number of regular seats to sell: R
- Number of discount seats to sell: D

Problem

		Price	Demand	Seats to Sell
JFK	Regular	617	100	
LAX	Discount	238	150	

Capacity 166

$$
\begin{aligned}
\text{Maximize} \quad & 617R + 238D \\
\text{Subject to} \quad & R + D \leq 166 \\
& R \leq 100, D \leq 150 \\
& R \geq 0, D \geq 0
\end{aligned}
$$

(Advanced) Objective

Objective: $\sum_{i\in\{R,D\}} p_i \text{ demand}_i(p_i)$

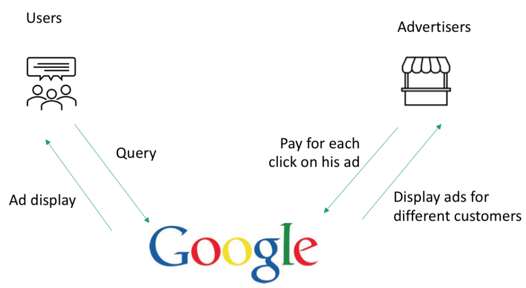
- We may further maximize the revenue by considering prices different from 617, 238.
- Demand forecasting

  Demand = f (price)

  - Demand for different prices can be forecasted using analytics tools, looking at historical data and incorporating models of human behavior
  - e.g., linear regression to obtain the relation: Demand = f(price)

案例分析3: online advertising

Eco-system



Users

- Different users may have different queries. Say one query from one user.

Query 1 ("4G LTE")	Query 2 ("largest LTE")	Query 3 ("best LTE network")

- The estimated number of query i is n_i

Query	Est. # of Requests
Q1 ("4G LTE")	140
Q2 ("largest LTE")	80
Q3 ("best LTE network")	80

- When viewing ad k, the probability that a user with query i will click on the ad is ρ_{ik}. **clicks per user rate**

Advertiser	Query 1 ("4G LTE")	Query 2 ("largest LTE")	Query 3 ("best LTE network")
AT&T	0.10	0.10	0.08
T-Mobile	0.10	0.15	0.10
Verizon	0.10	0.20	0.20

Advertisers

- The price that advertiser k wants to pay when his ad is clicked by a user with query i is p_{ik}. **price-per-click**
  - Maybe this is revenue earned from query i by advertiser k

Advertiser	Query 1 ("4G LTE")	Query 2 ("largest LTE")	Query 3 ("best LTE network")
AT&T	\$5	\$5	\$20
T-Mobile	\$10	\$5	\$20
Verizon	\$5	\$20	\$25

- Advertiser k has a budget of b_k

Advertiser	Budget
AT&T	\$170
T-Mobile	\$100
Verizon	\$160

Google

- Google's problem: How many times to display each ad for each query to maximize revenue.
- Decisions
  - x_{ik}: the number of query i for which Google displays ad k.
- How many users with query i clicking on ad k:

  $x_{ik}\ \rho_{ik}$
- How much google earns for all users with query i clicking on ad k:

  $x_{ik}\ \rho_{ik}\ p_{ik}$
- **Objective:** $\sum x_{ik}\ \rho_{ik}\ p_{ik}$
- **Decisions:**
  - x_{ik}, the number of query i for which Google displays ad k.
- **Objective:** $\sum x_{ik}\ \rho_{ik}\ p_{ik}$
- **Constraints:**
  1. Decision variables shall be positive.

     $0 \le x_{ik}$
  2. There are n_i users with query i.

     $\Sigma_k x_{ik} \le n_i$
  3. The price an advertiser paid cannot exceeds the budget.

     $\Sigma_i x_{ik}\ \rho_{ik}\ p_{ik} \le b_k$

§3 Optimization: How to solve?

1. optimization 的种类

1° Convex / nonconvex optimization (凸规划/非凸规划)

2° Constrained / unconstrained optimization (约束优化/无约束优化)

2° Convex v.s. nonconvex

1° Convex problem

- 例: linear optimization
- 可以在 polynomial time (多项式时间) 内有效解决 (find global optimal solution)
- Gradient descent (梯度下降法) (或其他 acceleration methods e.g. Newton method) can converge to global solution.

2° Non-convex problem

- 不可以在 polynomial time 内 find global optimal solution
- 可以使用 heuristics (试探法) to find local optimal solution