Assortment optimization

Problem statement:

Determine the optimal assortment of SKU's per store, Such that the Revenue per Store is maximized.

Revenue]

$$\text{Max } R_{\text{revenue}} \sum_{i=1}^{n \text{ (all SKU's)}} R_i(\text{Price}, \underbrace{\text{Nos of units}}_{\text{decision vector For optimization}}, \text{demand})$$

Discrete choice model.

Decision Vector Variable.

$$D = \begin{bmatrix} n_1 \\ n_2 \\ n_i \\ \vdots \\ n_j \end{bmatrix}$$

where n_i = (Nos of units of 'ith' SKU)

'i' → 'i'th SKU

$n_i \in (a, b)$ → $a = 0$, b = upper count of SKU

→ Thus we have an INTEGER optimization Problem at hand

[Demand] → Concept of utility.

→ ~~Util~~ The Probability of choosing a Particular SKU.

↓

Consider. UA1 to be the utility associated with SKU 'A' in store No '1'

Q] How do we map Probability to utility ?

→ What is utility a function of ?

→ Can we ~~determine~~ Estimate? this function approximately ?

Machine Learning ?

Mapping Probabilities to Utility.

Here we use a Logit model ::

We ~~get~~ have the following

For Store '1'

$$P_{A1} = \frac{\exp(U_{A1})}{1 + \sum_{j=A,B,C\ldots} \exp(U_{j1})}$$

RECAP → Logit

$$\text{Log}(\text{odds}) = \ln\left(\frac{p}{1-p}\right)$$

Now we have.

$$\ln\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \cdots \beta_k x_k$$
$$= \beta(X)$$

$$\therefore \frac{p}{1-p} = \exp(\beta(X))$$

$$\therefore p = \frac{\exp(\beta(x))}{1 + \exp(\beta(x))}$$

Model initialization

- Notes: We do not know the functional form for Utilities

initial guess:

$$\begin{bmatrix} U_{A1} \\ \vdots \\ U_{D1} \\ \vdots \end{bmatrix} = \{0\}$$

Zero Vector.

Set the utilities to zero

Example:

Assume store 1 has only 4 SKU's.

Then we have

$$U_{A1} = 0$$
$$U_{B1} = 0$$
$$U_{C1} = 0$$
$$U_{D1} = 0$$

initialization stage

→ This leads to

$$P_{A1} = \frac{\exp(0)}{1 + \sum_{j=A,B,C,D} \exp(U_{j1})} = \frac{1}{1+4} = 0.2$$

Thus

$$P_{A1} = P_{B1} = P_{C1} = P_{D1} = 0.2$$

iterate and update For Each store.

Step 4

ALL SKU'S

$$\begin{bmatrix} U^{t+1}_{A1} \\ \vdots \\ U^{t}_{D1} \\ \vdots \end{bmatrix} = U^{0}_{A1} + \overbrace{\log(SA1) - \log\left(\frac{PA1}{\sum_{j=A,B,C\ldots} PA_j}\right)}^{\Delta}$$

where SA1 = Observed market share of SKU A in store 1

We can determine SA1 as

$$SA1 = \frac{\text{Nos of Units of 'A' Sold in store 1}}{\text{Total units Sold in store 1}}$$

Explanation:

SA1 → observed share

$\frac{PA1}{\Sigma(PA's)}$ → Estimate of observed share.

error or Δ = $\log(SA1) - \text{Log}\left(\frac{PA1}{\Sigma PA's}\right)$

Error term is for utilities.

Step2 → Now check the 'Δ' if $\Delta <$ tolerance ~~Iteration~~ ~~Limit~~

or ITER > ITER Limit

↓

STOP.

Step3 update Probabilities.

↓

Loop Step 1 to Step 3 till Iteration Limit is reached.

OR error is Less than defined tolerance.

Estimate Functional Form for utilities.

We have after 'M' iterations

$$\begin{bmatrix} U_{A1}^{+M} \\ U_{D1}^{+M} \end{bmatrix} \quad \begin{bmatrix} X_{1}A,\ X_{2}A,\ X_{3}A \cdots \\ \\ \end{bmatrix}$$

feature vector (Row) For SKU 'A'

Assume

$$\underbrace{UA}_{\text{Utilities are Iteratively Computed}} = f\big(\underbrace{X_{1}A, X_{2}A \cdots X_{p}A}_{\text{features are known.}}\big)$$

Now using the 'Entire historical Data'

Say 12 monthly sales data.

or 'n' periods sales data.

Prepare a dataset as follows. For store 1

SKU	period	Feature vector For SKU. For periods.
UA	1	$X_{1}A$ $X_{2}A$ - - - .
UA	2	$X_{1}A$ $X_{2}A$ - - - - - .
UA	3	
	.	
UB	1	
UB	2	
	⋮	

Train Random Forest.

[U] on [X]

Let us Revisit the actual Business Problem Now.

Q] What is the Assortment of SKU's to be decided for store '1' For 'next Month'

To maximize Revenue, Given shelf Store space constraints!

Given Data For month 'p' + 1

$$[X]_{p+1}$$

Unknown: $[U]$

$$[X]_{p+1} \longrightarrow \begin{bmatrix} \text{Trained} \\ \text{Random} \\ \text{Forest} \end{bmatrix} \rightarrow [U]_{p+1}$$

$$[U]_{p+1} \rightarrow [P]_{p+1}$$ Compute Probabilities For ALL SKU's from Utilities.

↓

Optimization step / Function call.

Input [Revenue Function] $\longrightarrow$ $\sum_{i=1 \text{ to } n \text{ SKU's}} R\left[R_i = \text{Price}_i \times \text{Cnt}_i \times P_i \right]$

Price of SKU $_i$

Count of SKU $_i$

Prob. of Choosing SKU $_i$

Objective [Maximize Revenue]

Decision Variables [Vector of SKU Count]

Constraint [Total shelf space] $\rightarrow \sum_{i=1}^{n \text{ SKU's}} [\text{Vol of SKU } i \times \text{Cnt of SKU } i] <$ Shelf available