

# クーポン割り当て最適化実験

## 背景

クーポン A は 1000 円、クーポン B は 1500 円、クーポン C は 2000 円の割引を提供する。ここではこれまでクーポン 2 が全ユーザーに配布されていた状況を仮定する。

## 目的

クーポンの割り当てを最適化することにより、CV 数を最大化すること。

## 実験 1: 2 種類のクーポンを用いたコスト最小化

### 背景・目的

クーポン A とクーポン B のみを用いて、CV 数の棄損を抑えた上でのコスト最小化を行う。ここで浮いたコストが、より高い割引額のクーポンを配布する機会を生む。そのため、この実験を通して CV 数最大化に向けた改善の余地があるか確認する。

### 検討ロジック

検討したロジックは大きく次の 2 段階に分かれる。

1. 機械学習モデルにより、CVR とポイント利用率 (PUR; Point Utilization Rate) を予測する。
2. 予測値を用いて CV 数を最大化する数理最適化問題をソルバーで解き、ユーザーへの最適なクーポン配布を行う。

まず、今回解きたい問題を以下のように定式化した。

$$\begin{aligned}
& \text{Minimize} && \sum_{s \in S} \sum_{a \in A} x_{s,a} N_s r_{s,a} p_{s,a} C_a \\
& \text{s.t.} && x_{s,a} \in [0, 1] \\
& && \sum_{a \in A} x_{s,a} = 1, \forall s \in S \\
& && \sum_{s \in S} \sum_{a \in A} x_{s,a} N_s r_{s,a} \geq R
\end{aligned}$$

表記は以下の通りである。

- $S$ : ユーザーセグメント集合
- $A$ : クーポン集合
- $x_{s,a}$ : セグメント  $s$  へのクーポン  $a$  の配布率
- $r_{s,a}$ : セグメント  $s$  にクーポン  $a$  を配布した時に CV する確率 (CVR)
- $p_{s,a}$ : セグメント  $s$  にクーポン  $a$  を配布して CV した際にクーポンを利用する確率 (PUR)
- $N_s$ : セグメント  $s$  に属するユーザー数
- $C_a$ : クーポン  $a$  を使って CV された際に発生するコスト
- $R$ : 許容できる CV 数の下限

この最小化問題を解くことで、セグメント  $s$  にクーポン  $a$  を何%配布するかを決めることができる。最小化問題を解く上で、 $r_{s,a}$  と  $p_{s,a}$  は未知なため、機械学習モデルから得られる予測値を用いた。

最適な  $x_{s,a}$  を得た後のユーザーに対するクーポン割り当ては、配布割合を用いて確率的に行なった。確率的の方策にしておくことで、将来的にログデータを用いたモデルのエンハンスを行いやすくすることを狙った。

## 実験方法

ダミーデータを生成し、64%を学習データ、16%を検証データ、20%をテストデータとした。CV したかと CV した上でポイント利用したかの両方において分布に偏りを生じさせなかったため、データの分割には `iterstrat` を用いた。

CVR 予測モデルは、CV したかを目的変数とし、LightGBM を Binary Logloss で学習させることで得た。PUR 予測モデルは、CV した上でクーポンを利用したかを目的変数とし、LightGBM を Binary Logloss で学習させることで得た。

ユーザーセグメントは、次の 2 段階で行なった。

1. CVR 予測モデルによって得られたクーポン A 付与時 CVR をもとに、pandas の `qcut` で 10 分割。
2. 1.で得られたセグメントごとに、クーポン B 付与時 CVR をもとに pandas の `qcut` で 10 分割。

セグメントごとのユーザー数に偏りが出ないこと、なるべく CVR が近いユーザーを同じグループにまとめようことを狙った。

予測値を用いた数理最適化は、`pulp` を用いて解いた。最適な  $x_{s,a}$  を得た後のユーザーに対するクーポン割り当ては、配布割合を用いて確率的に行なった。

最適な割り当てを行なった際に生じる CV 数とコストの推定は、次のように行なった。

1. 最適な割り当てを行なった場合のクーポンと、ログ上のクーポンが一致しているデータを抽出。
2. 1.のデータを用いて、クーポンごとに CVR と PUR を計算。
3. クーポンごとの付与人数に 2.で求めた CVR をかけることで推定 CV 数を計算。
4. 3.で求めた推定 CV 数に PUR とクーポンの割引額をかけることで推定コストを計算。

## 結果

### CVR 予測モデルと PUR 予測モデル

CVR 予測モデルの学習曲線を図1、PUR予測モデルの学習曲線を図2に示す。CVR予測モデルは損失が下がっていく傾向が見られたが、PUR予測モデルは損失が上がる傾向が見られた。PUR予測モデルに利用できるデータはCVしたデータのみであり、サンプル数が少ないためCVR予測モデルよりも学習が難しかった可能性や、説明変数と目的変数の間で関連が弱かった可能性が考えられる。

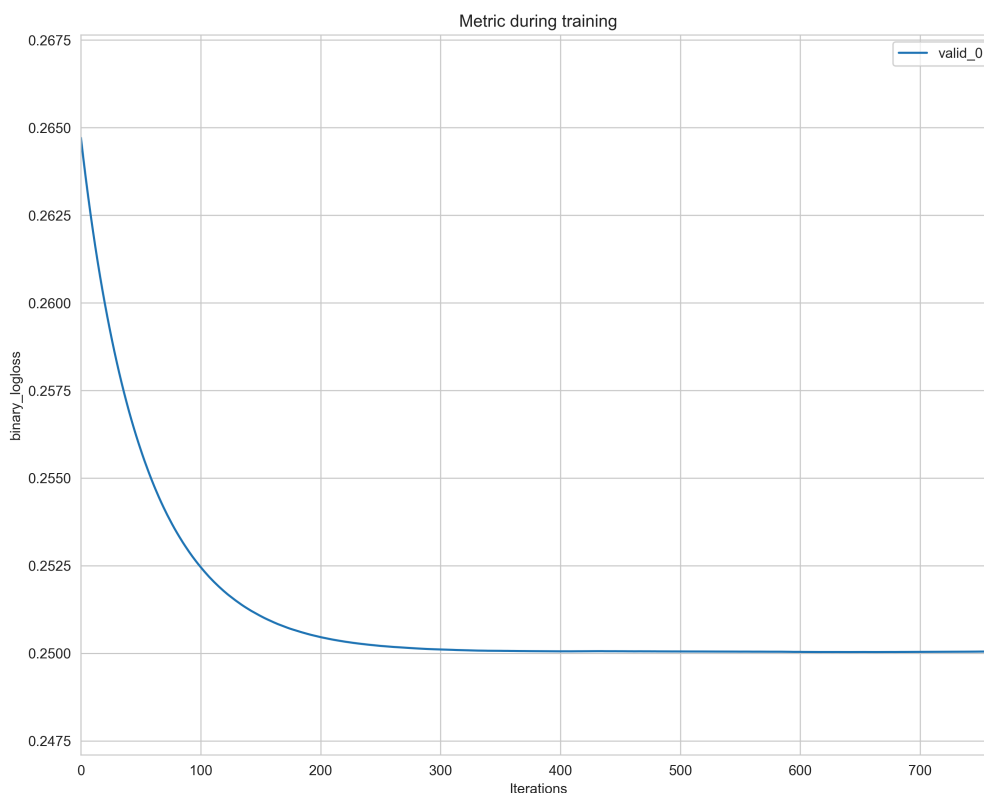


図1: CVR予測モデルの学習曲線

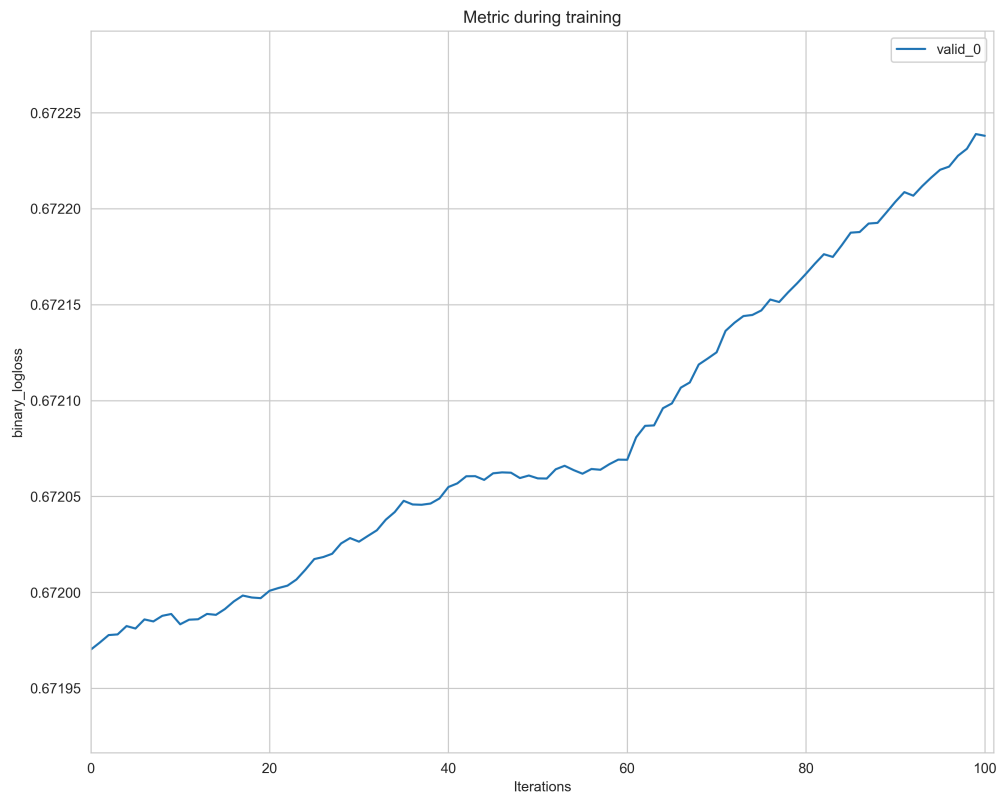


図2: PUR予測モデルの学習曲線

機械学習モデルの確率値が適切にキャリブレーションされているかを確認するため、横軸に予測CVR、縦軸に実測CVRをとった散布図を図3と図4に示す。図3はクーポンA、図4はクーポンBの結果である。予測値と実測値に大きな乖離はないことが確認できる。

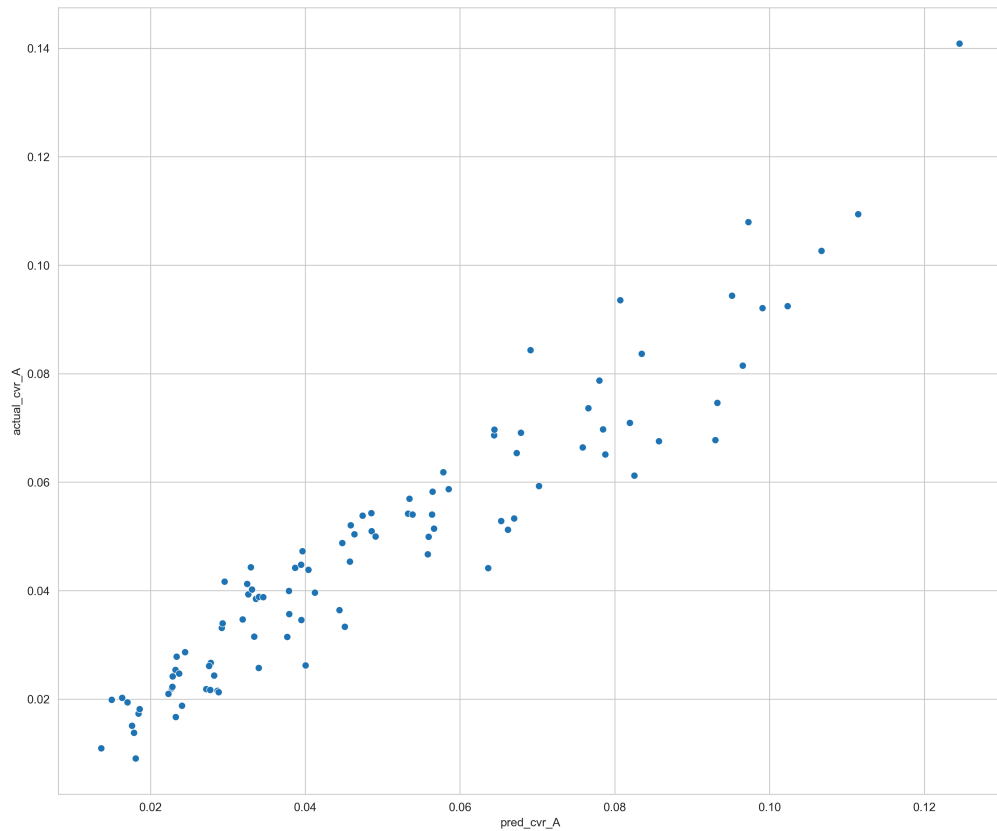


図3: クーポンAに対する予測CVRと実測CVR

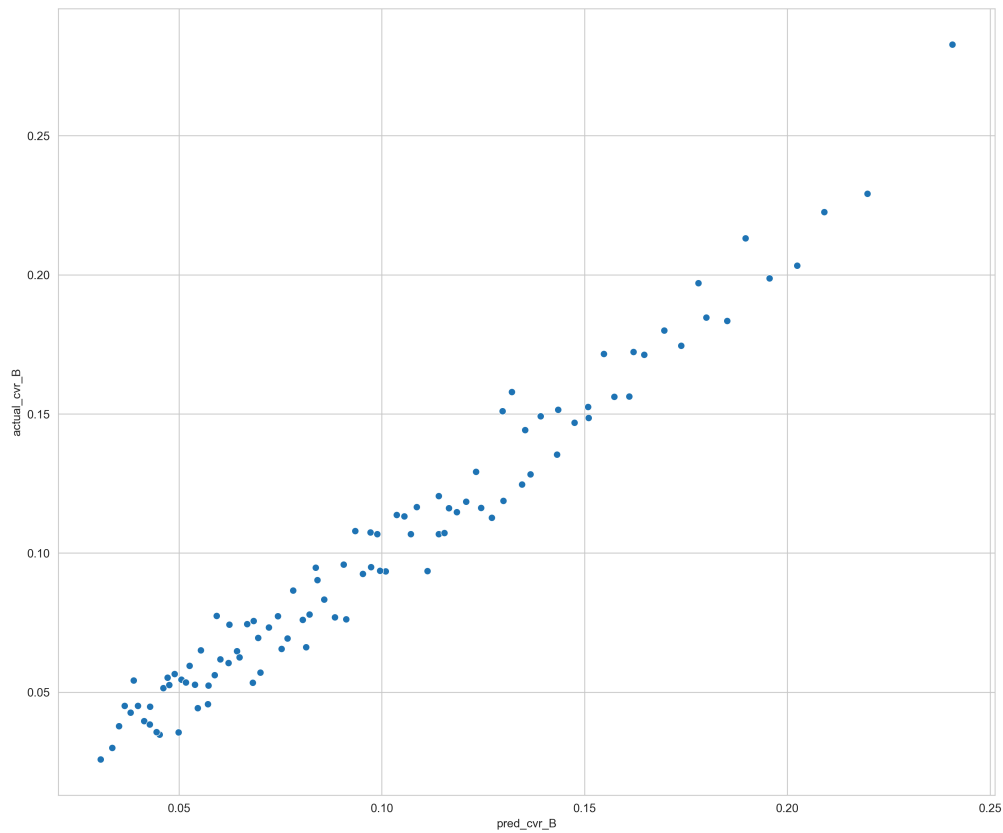


図4: クーポンBに対する予測CVRと実測CVR

PURにおける同様の散布図を、図5と図6に示す。図5がクーポンA、図6がクーポンBに対する結果である。PURは予測値が0.6付近に集中している一方、実測値は0.4から0.8程度までばらついている傾向が見える。PUR予測モデルの学習がうまくいかなかったため、予測値と実測値に大きな乖離が生まれたと考えられる。

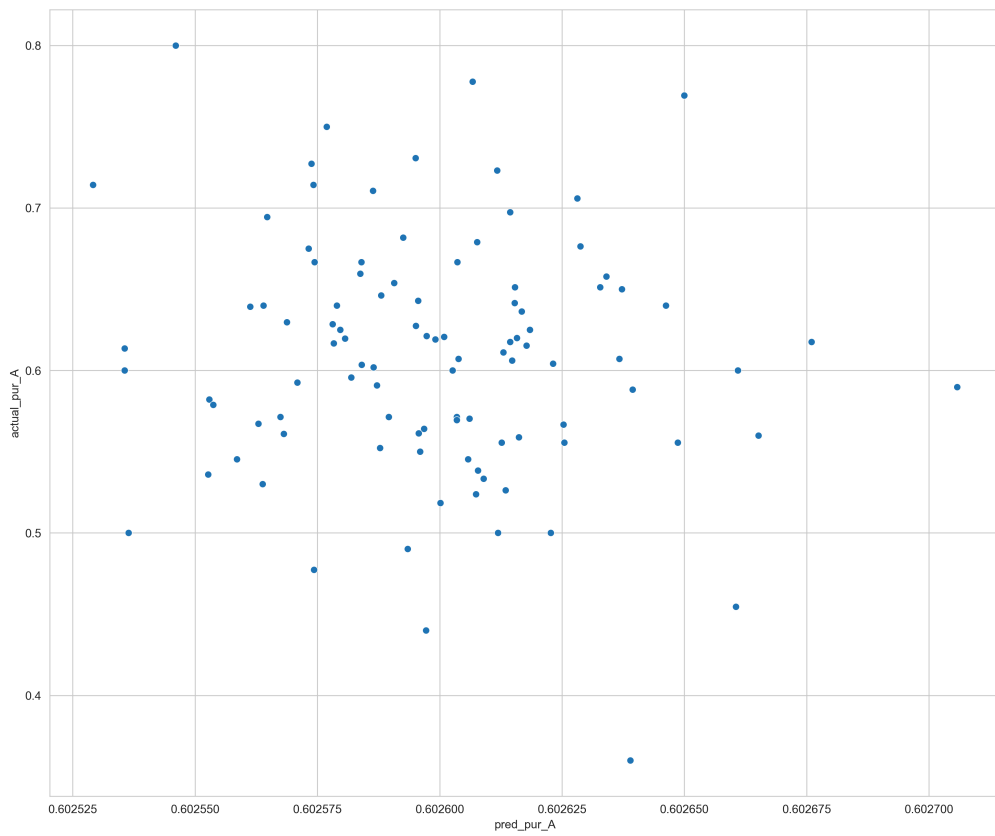


図3: クーポンAに対する予測PURと実測PUR

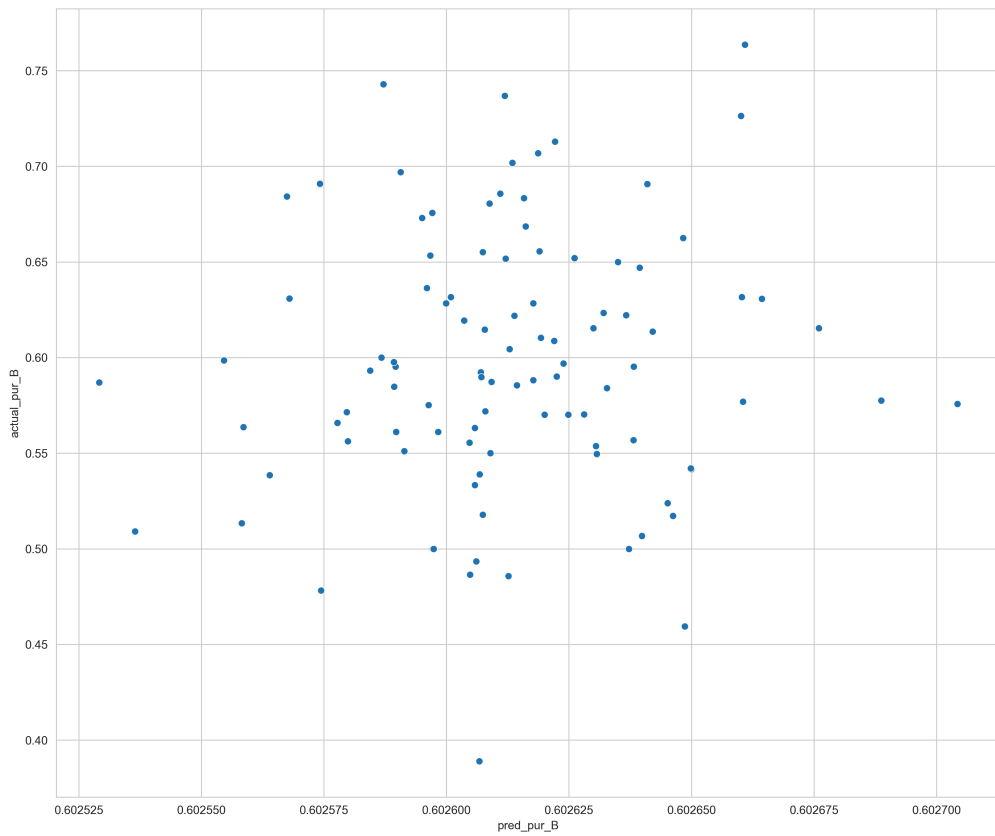


図4: クーポンBに対する予測PURと実測PUR

コスト最小化を狙ったクーポン割り当て

## 実験 2: 3 種類のクーポンを用いた CV 数最大化

背景・目的

検討ロジック

実験方法

結果

まとめ