

# スーパーマーケットにおける 商品の売上個数の予測

中央大学大学院 理工学研究科 経営システム工学専攻

生田目研究室

齋藤 烈也

1. 研究背景
2. 目的
3. 基礎集計
4. 分析
  1. Xgboost
  2. 説明変数
5. 分析結果
6. まとめ

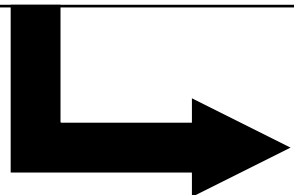
## 学術的背景

スーパーマーケットの店舗でのマーチャンダイジングにおいて、発注とそれに伴う需要予測は最も重要なものの一つとされている。 [木下 11]

なぜなら、発注と需要予測の巧拙で店舗の売上や損失が大きく左右されるからである。

## 実際起きている問題

- ・ 前日と同じ量が無条件で機械的に発注するガラガラ発注
- ・ 廃棄を減らすように発注量を大幅に少なくする



損失をコントロールするという観点からは必ずしも適切な発注業務がなされているとは限らない現状

## 目的

スーパーマーケットにおける商品の売上数量を予測する

## 使用データ

使用データ	スーパーマーケットのID-POSデータ
対象年	2014年4月1日～2016年3月31日
対象人数	20,297人（会員のみ）
商品数	106,749種
レシート数	1,223,961枚

※このうち2014/4/1～2016/2/28までのデータを訓練データとして、  
2016年3月中の商品売上数を予測する

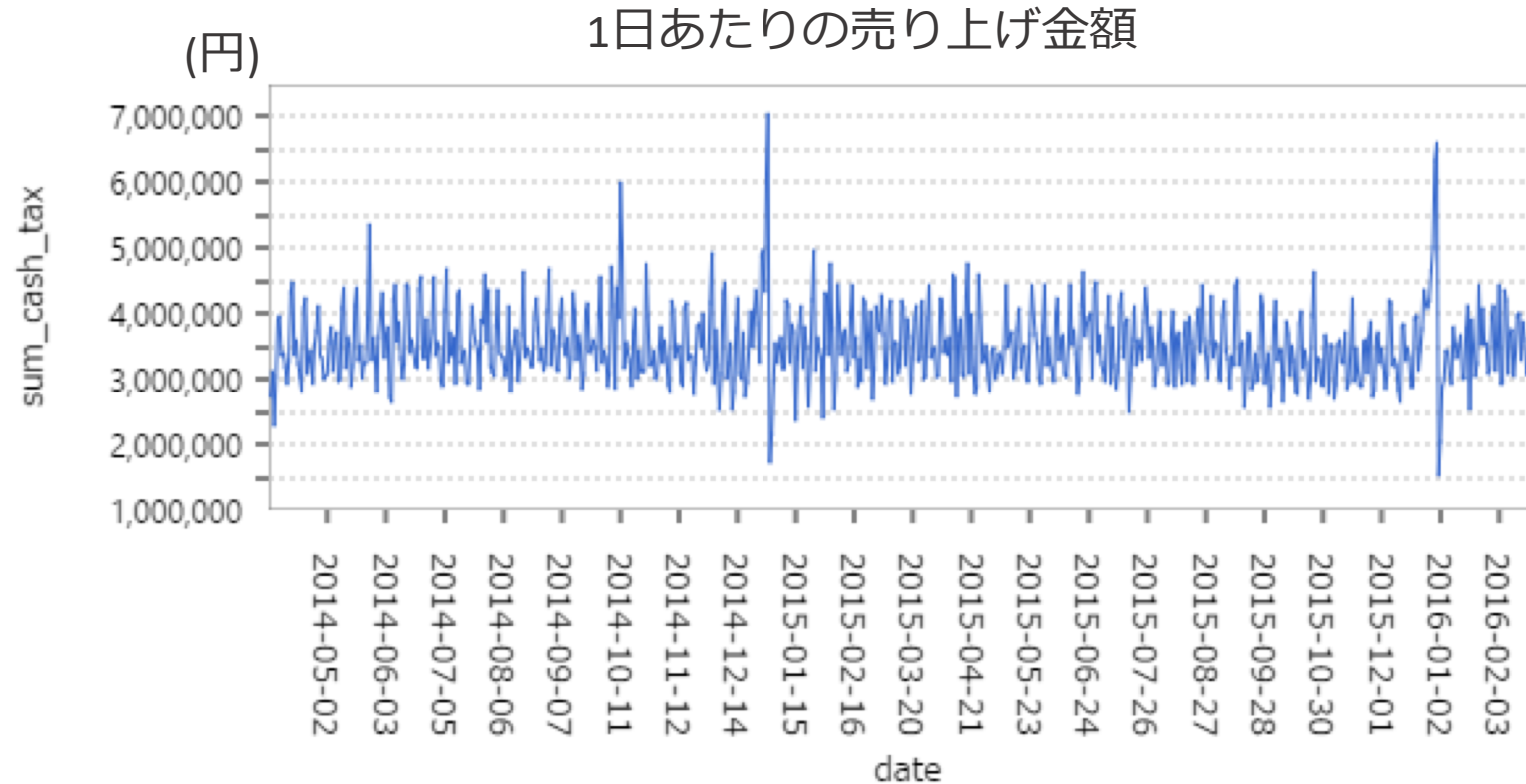
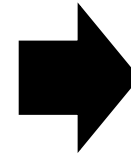
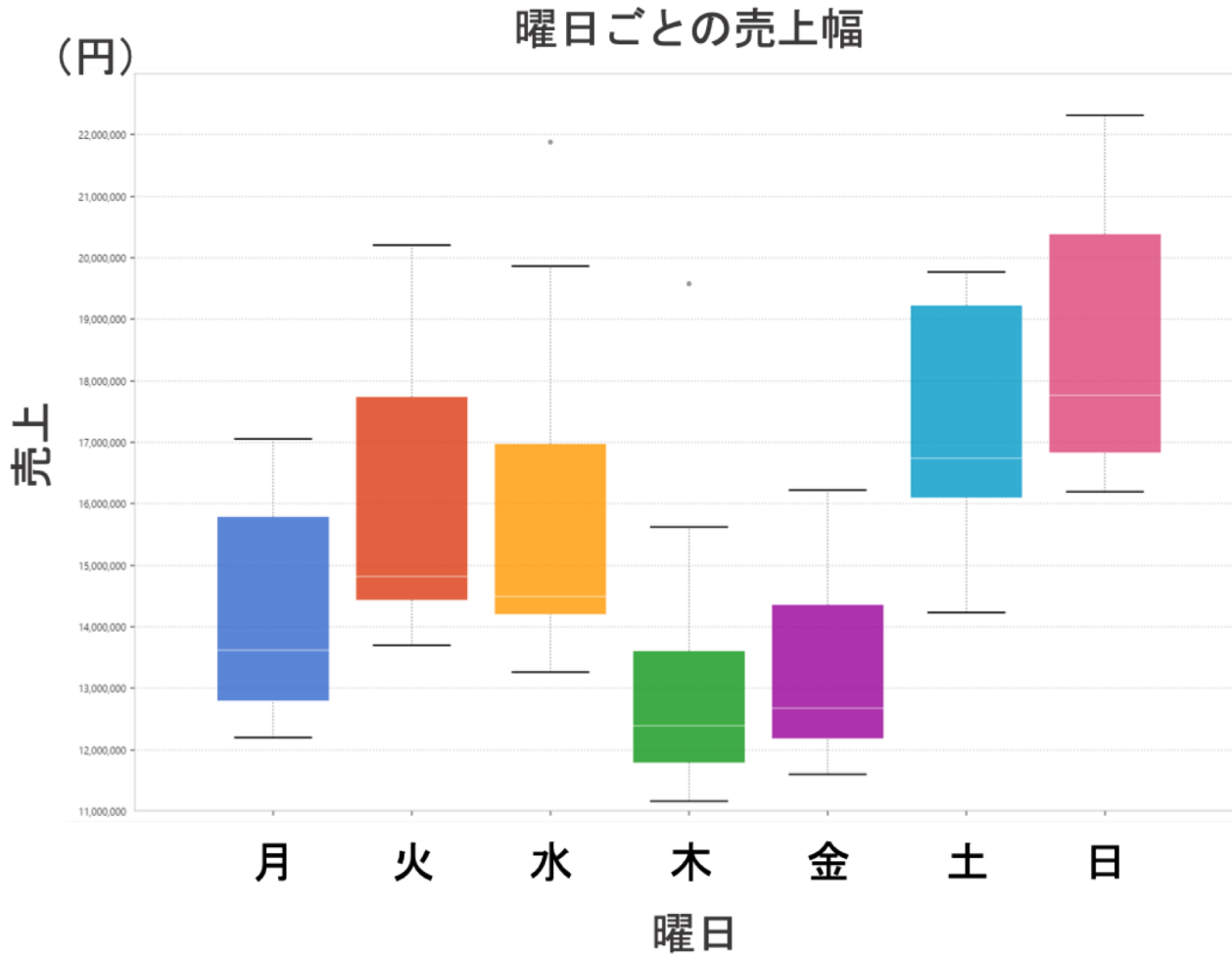


図1 1日当たりの売り上げ金額

➡ 年末年始に大きな波があり，全体の売り上げ傾向としてはほぼ一定である



土日の売上が高く,  
月木金と売上が低い

図2 曜日ごとの売上幅

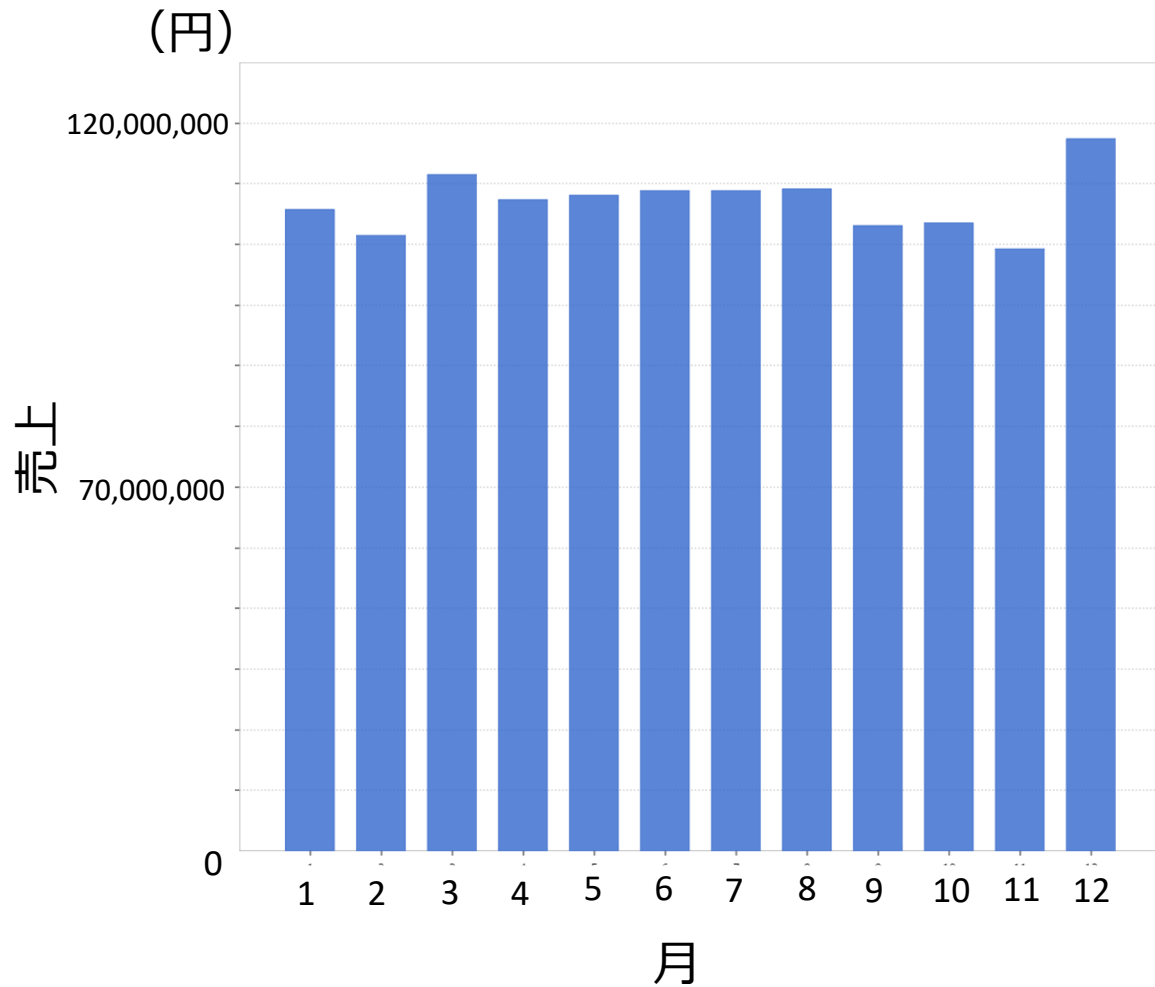


図3 月ごとの売上 (2015年)

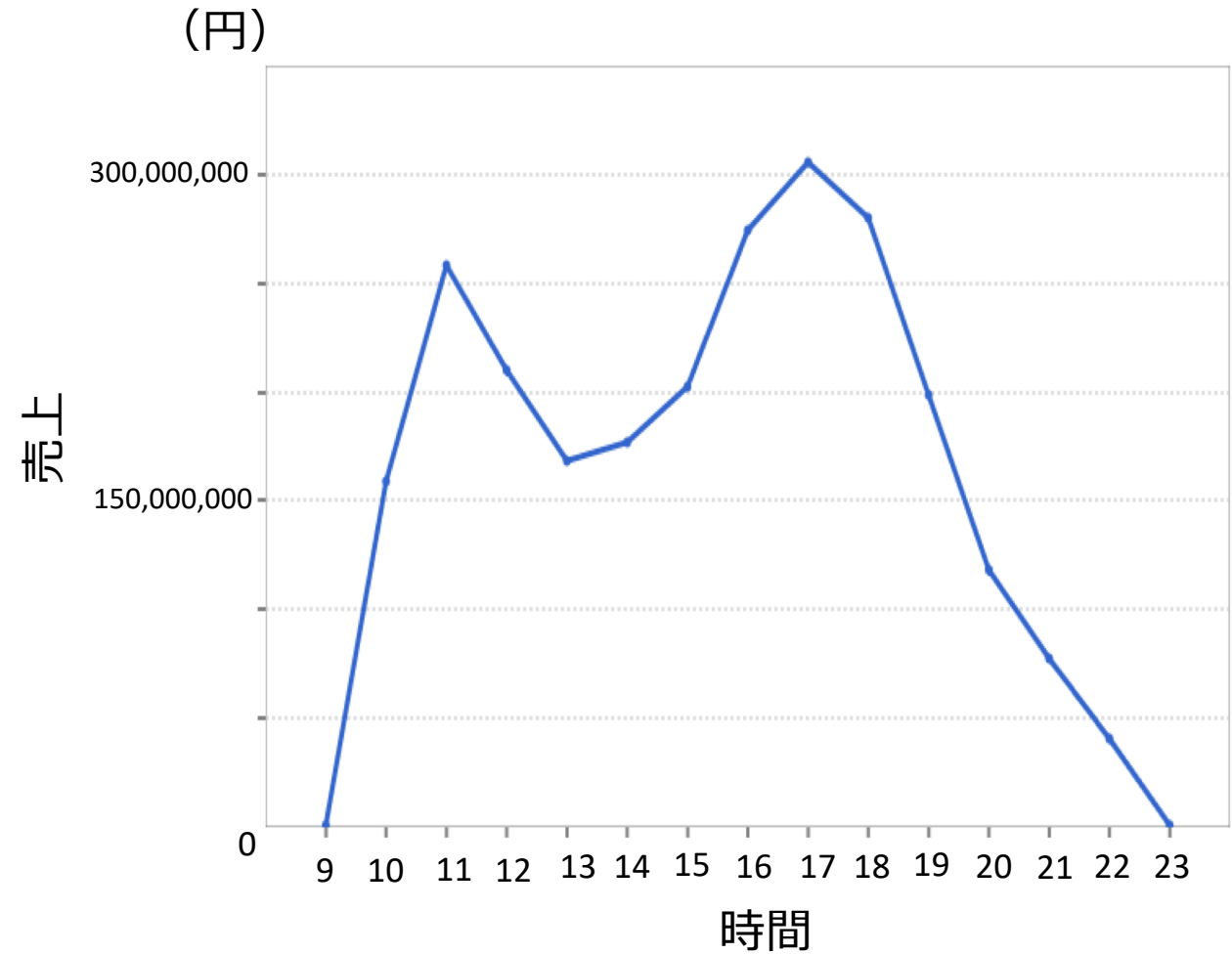


図4 時間ごとの総売り上げ

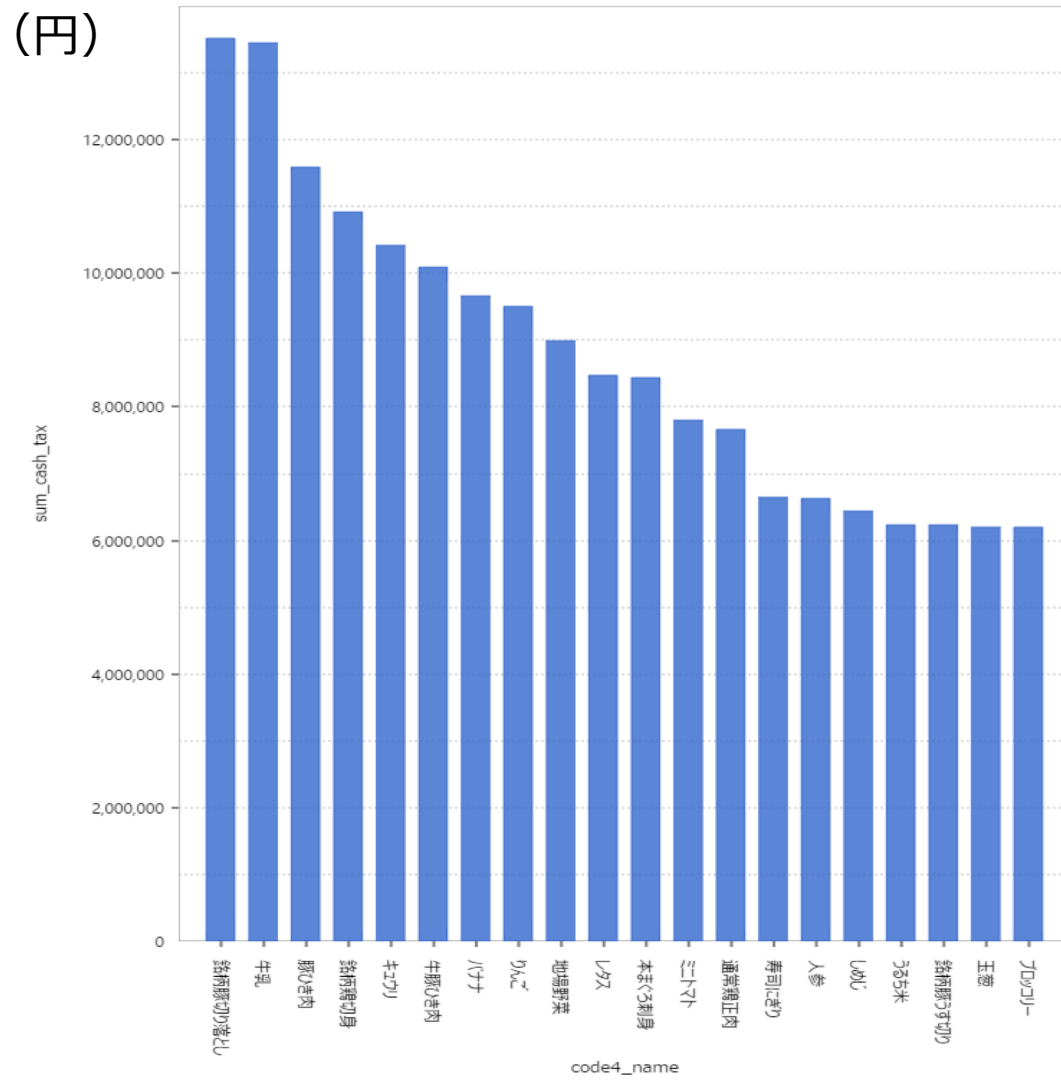


図5 売上上位商品

表1 売上上位商品

順位	商品名
1	銘柄豚切り落とし
2	牛乳
3	豚ひき肉
4	銘柄鶏切り身
5	きゅうり



今回の予測にはXgboostを使用 [Chen 16]

Xgboost・・・勾配ブースティングとも呼ばれる手法で、**弱学習学習器を逐次的に構築する手法**

ランダムフォレストとの違い

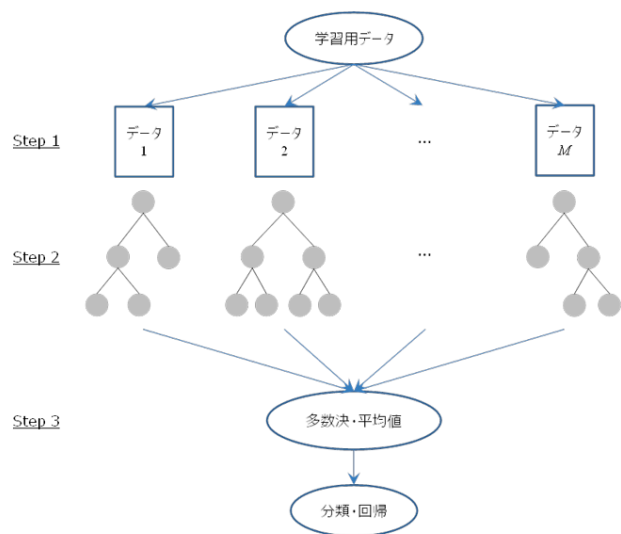


図6 ランダムフォレスト (参照: 日経リサーチ)

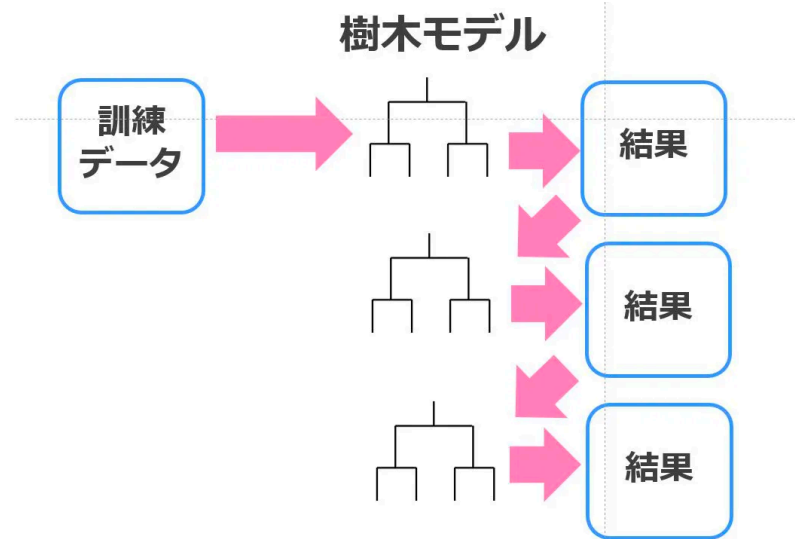


図7 Xgboost (参照: スタビジ)

ある入力 $x_i$ に対して $y_i$ を予測することを考える.

1. 1つのモデル $f_1$ を作成する. このモデルの出力を $y_i$ に対する予想値 $\hat{y}_i^1$ とする

$$\hat{y}_i^1 = f_1(x_i) \quad (1)$$

モデルの予想した結果と実際の $y_i$ の差が生じる

$$err_i^1 = y_i - \hat{y}_i^1 \quad (2)$$

2.  $err_i^1$ を出力するようなモデル $f_2$ を作成

$$\hat{y}_i^2 = \hat{y}_i^1 + f_2(x_i) = f_1(x_i) + f_2(x_i) \quad (3)$$

3.  $K$ 回繰り返すと予測値は次のようになる

$$\hat{y}_i^{(K)} = \sum_{k=1}^K f_k(x_i) \quad (4)$$

損失関数 $l(a, b)$ を導入する

$$\begin{aligned} & \min_{f_t} \sum_{i=1}^n l(y_i, \hat{y}_i^{(k)}) \\ \Leftrightarrow & \min_{f_t} \sum_{i=1}^n l\left(y_i, \sum_{k=1}^t f_k(x_i)\right) \\ \Leftrightarrow & \min_{f_t} \sum_{i=1}^n l\left(y_i, \sum_{k=1}^{t-1} f_k(x_i) + f_t(x_i)\right) \\ \Leftrightarrow & \min_{f_t} \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)\right) \end{aligned}$$

このままでは過学習を起こしてしまうので罰則項を追加する.

$$\Omega(f_t) = \gamma T + \frac{1}{2} \lambda \|w\|^2 \quad (5)$$

上記の罰則項を損失関数に加えたものを  $\mathcal{L}^{(t)}(f_t)$  とする.

$$\begin{aligned} \min_{f_t} \mathcal{L}^{(t)}(f_t) &= \min_{f_t} \sum_{i=1} l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \\ &= \min_{f_t} \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \gamma T + \frac{1}{2} \lambda \|w\|^2 \end{aligned}$$

ここで  $T$  は決定木を構築した時の最終ノードの数であり,  $\gamma$ ,  $\lambda$  はハイパーパラメータである. また  $w$  は決定木  $f_t$  が返すことのできるベクトルである.

## Xgboostのアルゴリズムの導出

1. 目的関数 $\mathcal{L}^{(t)}(f_t)$ の第1項、 $l$ を $f_t$ に関して0の周りで2次のテーラー展開を行う.

$$\begin{aligned}\min_{f_t} \mathcal{L}^{(t)}(f_t) &= \min_{f_t} \sum_{i=1}^n l(y_i, \hat{y}_i^{(t-1)} + f_t(x_i)) + \Omega(f_t) \\ &\approx \min_{f_t} \sum_{i=1}^n \left( l(y_i, \hat{y}_i^{(t-1)}) + g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right) + \Omega(f_t) \\ &\propto \min_{f_t} \sum_{i=1}^n (g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i)) + \Omega(f_t)\end{aligned}$$

ここで,  $g_i = \partial_{\hat{y}_i^{(t-1)}} l(y_i, \hat{y}_i^{(t-1)})$ ,  $h_i = \partial_{\hat{y}_i^{(t-1)}}^2 l(y_i, \hat{y}_i^{(t-1)})$

ここで,  $\mathcal{L}^{(t)}(f_t)$ を以下のように書き直す.

$$\begin{aligned}\min_{f_t} \tilde{\mathcal{L}}^{(t)}(f_t) &= \min_{f_t} \sum_{i=1}^n \left( g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right) + \Omega(f_t) \\ &= \min_{f_t} \sum_{i=1}^n \left( g_i f_t(x_i) + \frac{1}{2} h_i f_t^2(x_i) \right) + \gamma T + \frac{1}{2} \lambda \|w\|^2 \\ &= \min_{f_t} \sum_{j=1}^T \left( \sum_{i \in I_j} g_i f_t(x_i) + \frac{1}{2} \sum_{i \in I_j} h_i f_t^2(x_i) \right) + \gamma T + \frac{1}{2} \lambda \sum_{k=1}^T w_j^2 \\ &= \min_{f_t} \sum_{j=1}^T \left( \sum_{i \in I_j} g_i w_j + \frac{1}{2} \sum_{i \in I_j} h_i w_j^2 \right) + \gamma T + \frac{1}{2} \lambda \sum_{k=1}^T w_j^2 \\ &= \min_{f_t} \sum_{j=1}^T \left( \sum_{i \in I_j} g_i w_j + \frac{1}{2} \sum_{i \in I_j} h_i w_j^2 + \frac{1}{2} \lambda w_j^2 \right) + \gamma T \\ &= \min_{f_t} \sum_{j=1}^T \left( \sum_{i \in I_j} g_i w_j + \frac{1}{2} \left( \sum_{i \in I_j} h_i + \lambda \right) w_j^2 \right) + \gamma T\end{aligned}$$

$\mathcal{L}^{(t)}(f_t)$ を  $w_j$  で微分したものを0とおいて解くと最適解  $w_j^*$ , **つまり目的関数を最小にする  $t$  個目の決定木の最終ノード  $j$  が返すべき結果は**

$$w_j^* = -\frac{\sum_{i \in I_j} g_i}{\sum_{i \in I_j} h_i + \lambda} \quad (6)$$

ここから, 勾配情報を利用していどのように説明変数を分岐させるか考えていく.

2.  $w_j^*$ を $\mathcal{L}^{(t)}(f_t)$ に代入する

$$\mathcal{L}^{(t)}(f_t) = -\frac{1}{2} \sum_{j=1}^T \frac{(\sum_{i \in I_j} g_i)^2}{\sum_{i \in I_j} h_i + \lambda} + \gamma T \quad (7)$$

ノード  $s$  を分岐する前の目的関数を

$$\tilde{\mathcal{L}}_{\text{before}}^{(t)} = -\frac{1}{2} \sum_{j \neq s}^T \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_{i+\lambda}} - \frac{1}{2} \frac{\left(\sum_{i \in I_s} g_i\right)^2}{\sum_{i \in I_s} h_{i+\lambda}} + \gamma T$$

ノード  $r$  を分岐した後左側のノードに入るデータの集合を  $I_L$ , 右側のノードに入るデータの集合を  $I_R$  としたときの ( $I_s = I_L \cup I_R$ ) 分岐後の目的関数を

$$\tilde{\mathcal{L}}_{\text{after}}^{(t)} = -\frac{1}{2} \sum_{j \neq s}^T \frac{\left(\sum_{i \in I_j} g_i\right)^2}{\sum_{i \in I_j} h_{i+\lambda}} - \frac{1}{2} \frac{\left(\sum_{i \in I_L} g_i\right)^2}{\sum_{i \in I_L} h_{i+\lambda}} - \frac{1}{2} \frac{\left(\sum_{i \in I_R} g_i\right)^2}{\sum_{i \in I_R} h_{i+\lambda}} + \gamma T$$

最も目的関数  $\mathcal{L}^{(t)}(f_t)$  の値が小さくなるように分岐をするためには、**分岐前と分岐後の目的関数の差をとって、もっとも差が大きくなるところで分岐を行えば良い**ということになる。



$$\begin{aligned}\tilde{\mathcal{L}}_{\text{before}}^{(t)} - \tilde{\mathcal{L}}_{\text{after}}^{(t)} &= \frac{1}{2} \frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{1}{2} \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{1}{2} \frac{(\sum_{i \in I_S} g_i)^2}{\sum_{i \in I_S} h_i + \lambda} \\ &= \frac{1}{2} \left( \frac{(\sum_{i \in I_L} g_i)^2}{\sum_{i \in I_L} h_i + \lambda} + \frac{(\sum_{i \in I_R} g_i)^2}{\sum_{i \in I_R} h_i + \lambda} - \frac{(\sum_{i \in I_S} g_i)^2}{\sum_{i \in I_S} h_i + \lambda} \right)\end{aligned}$$

すべての説明変数  $x_i = (x_{i1}, \dots, x_{im})^T$  のあらゆる分割の候補のうち, 最も  $\tilde{\mathcal{L}}_{\text{before}}^{(t)} - \tilde{\mathcal{L}}_{\text{after}}^{(t)}$  を大きくする候補を 1 つ選んでそこで分岐を行うということである.

以上のような分析手法を用いて分析を商品数予測を行う.

# 分析（説明変数）

目的変数：2016年3月の各商品売上数（0～400に丸めてある）

説明変数	概要
観測開始月からの経過月	2014年4月を1として該当月の番号
商品番号	10桁34806種の番号
商品分類番号	1133種の数字
商品大分類	1～9の数字
商品中分類	99., 11., 12., 13., 14., 21., 22., 23., 31., 32., 33., 34., 35., 41., 42., 43., 44., 45., 51., 52., 53., 61., 62., 63., 71., 88.
○ヶ月前各商品売上数	1, 2, 3, 6, 12ヶ月前の商品売上数
○ヶ月前平均商品売上数	1, 2, 3, 6, 12ヶ月前の平均商品売上数
○ヶ月前各商品カテゴリ売上数	1, 2, 3, 6, 12ヶ月前の各商品分類売上数
○ヶ月前商品単価上昇率	$(\text{○ヶ月前各商品平均単価} - \text{当月各商品平均単価}) / \text{当月各商品平均単価}$
月, 日数	12ヶ月, 各月の日数
最終販売月, 売り始めた月	最後に販売していた月, 売り始めた月

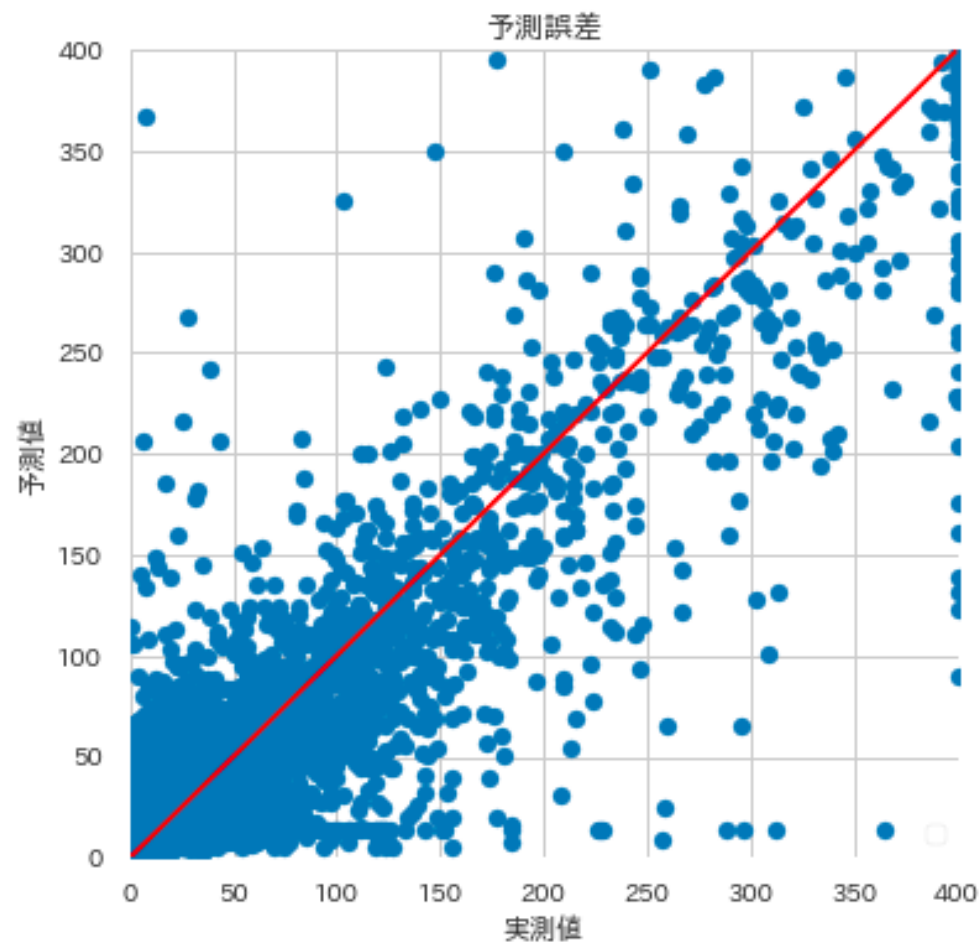


図8 予測誤差結果

表3 RMSEの結果

<b>RMSE</b>	<b>23.50569</b>
-------------	-----------------

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{k=1}^n (f_i - y_i)^2}$$

$f_i$  : 予測値,  $y_i$  : 実測値,  $n$  : サンプル数

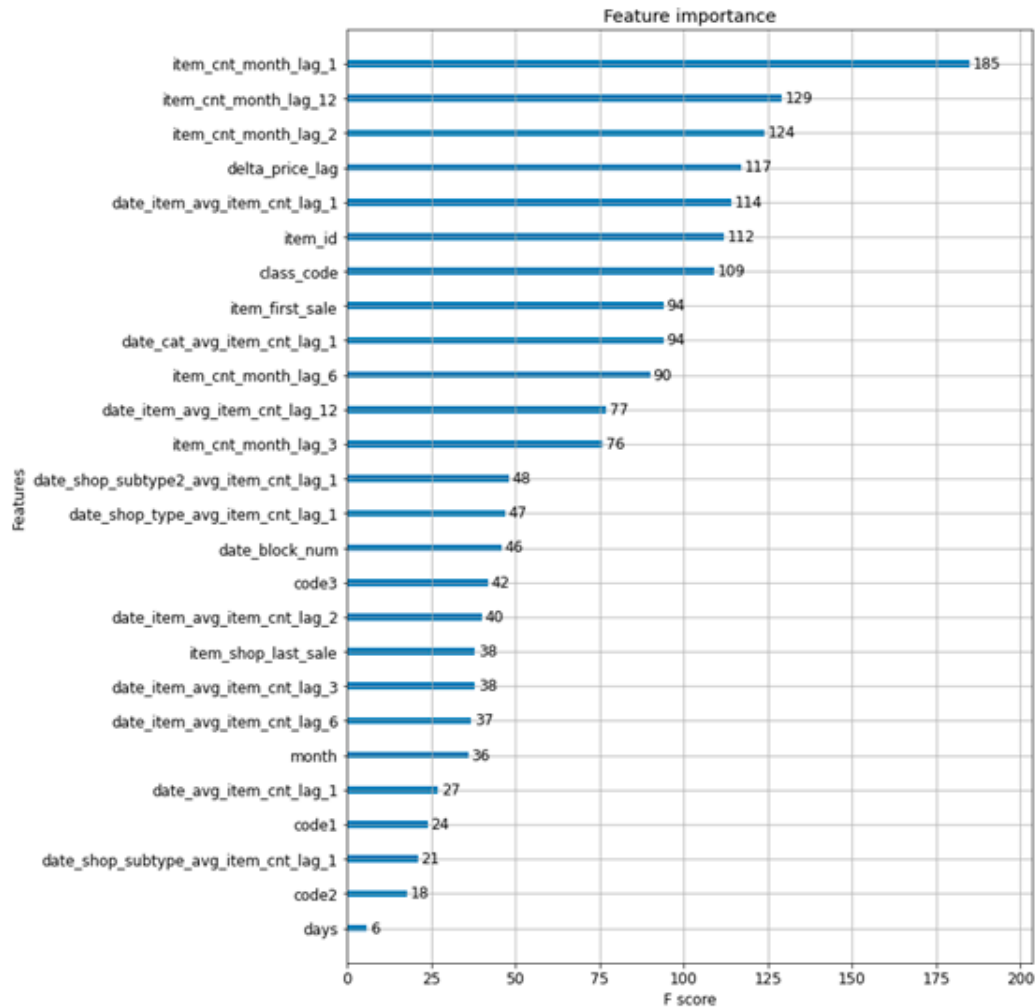
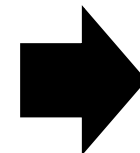


図9 変数重要度

表2 変数重要度順位

重要度順位	特徴量
1	先月商品売上数
2	1年前商品売上数
3	先々月商品売上数
4	前販売期販売単価上昇率
5	先月平均商品売上数
6	商品番号
7	カテゴリ番号
8	最初に売られた月
9	先月前各商品カテゴリ売上数
10	半年前商品売上数



商品の周期変動をとらえている変数の重要度が高い

## まとめ

- ・機械学習を用いて商品売上数の予測を行った。集計より、スーパーマーケットの特性上商品数の多さや季節性、周期性を考慮して特徴エンジニアリングを行った。その結果、変数重要度の高いものに周期性などを考慮した変数が多く見られた。
- ・売上個数の他にも、商品分類等にも着目し、説明変数として加えることで商品の潜在カテゴリ的要素にも付随して分析を行うことができた。

## 今後の方針

今回は商品ベースでの商品数予測に従事したが、ユーザベースでも商品数予測を行うことでユーザ属性等も考慮したマーチャンダイジングが可能になると考えられる。

- [ 1 ] コンビニエンスストアの知識 (木下安司, 2011)
- [ 2 ] XGBoost: A Scalable Tree Boosting System (Tianqi Chen and Carlos Guestrin, 2016)
- [ 3 ] POSデータに基づく欠品時の顧客購買行動を考慮した小売店舗の購買シミュレーション (松村直樹, 和泉潔, 山田健太, 2015)
- [ 4 ] 日常購買行動に関する大規模データの融合による顧客行動予測システム  
実サービス支援のためのカテゴリマイニング技術 (石垣司, 竹中毅, 本村陽一, 2011)
- [ 5 ] XgboostでKaggleの自転車需要予測 (<https://hampen2929.com/xgboost-kaggle-bike/724>)
- [ 6 ] 顧客行動の多様性変数を利用した購買行動の予測 (新美潤一郎, 星野崇宏, 2017)
- [ 7 ] Feature Engineering, Xgboost ( <https://www.kaggle.com/dlarionov/feature-engineering-xgboost> )