

小売業のポジショニングマップ

原田玲央

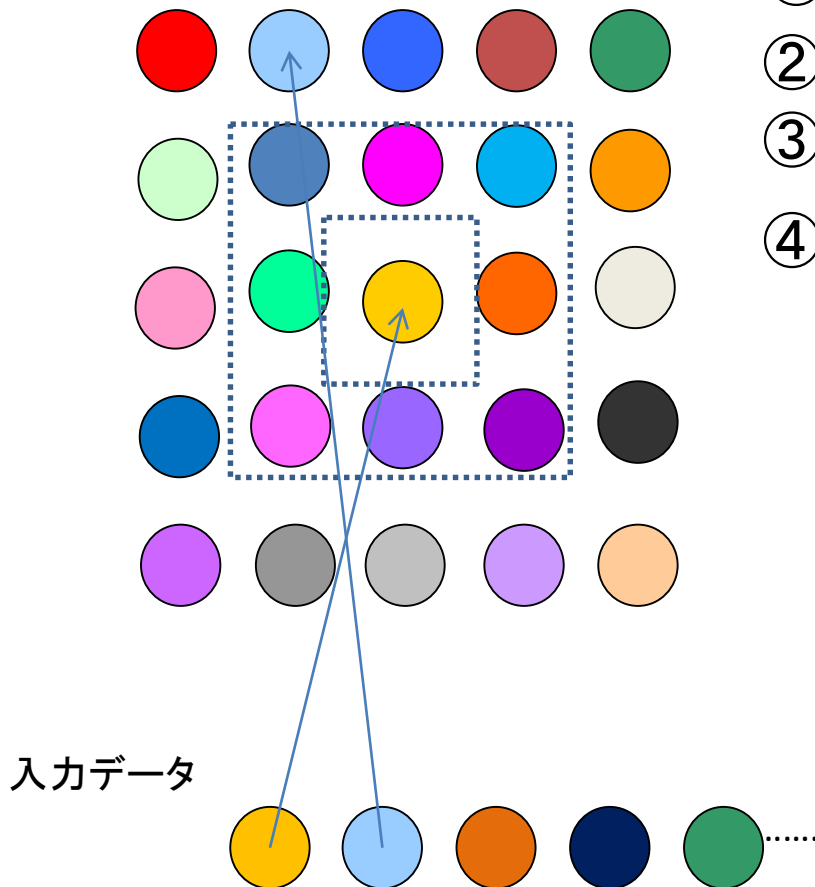
概要

- SOMのアルゴリズムをしっかりと理解する
- SOMとそれ以外の多変量解析の違いを理解する。
- 演習1＋解説
- SOMで表現できない情報を決定木で補う
- 演習2＋解説
- まとめ・感想
- 宿題

SOMのアルゴリズム

- ①入力データ1つを選択
- ②入力データと最も似た色を探す
- ③周囲の色を入力データに近いする
- ④①～③を繰り返す

これを入力データの数だけ行い学習1回となる



- この学習を何回も繰り返すとマップが完成する

スタート

線形空間との相違

- 多次元データを低次元に圧縮しマップにする方法はSOMだけではない。

→主成分分析、因子分析、多次元尺度法

(これらの手法はSOMを含め多変量解析の分野にあたる。)

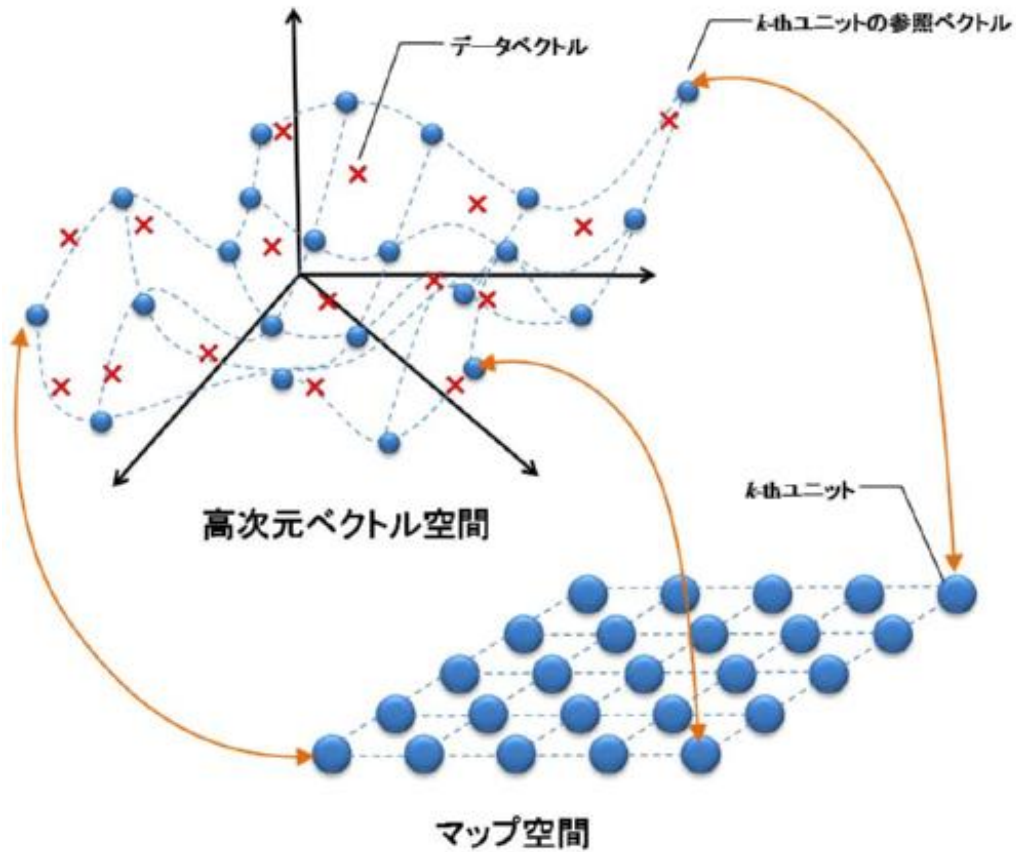
- 主成分分析、因子分析、判別分析、多次元尺度法

→線形写像を行いマップがユークリッド空間で描かれる

- SOM

→線形写像を行いマップがユークリッド空間で描かれない

写像を行うとは？



SOM以外の多変量解析

- 主成分分析 → 情報を縮約する
- 因子分析 → 共通因子を見つける
- 判別分析 → どちらのグループに属するか推測する関数を見つける
- 多次元尺度法 → 距離データを低次元に配置する手法

ユークリッド空間

- ユークリッド空間とは距離が通用する空間
- SOM以外のマップは通常的地図と同じように解釈できる。
- SOMの隣り合うユニットは離れたユニットより類似度が高い。
 - n個離れてたら類似度が $1/n$ になるわけではない。

主成分分析とSOMの違い

ココのデメリットを解決していく

	主成分分析	SOM
記述の対象 (マップに描写される 情報)	変数(←主成分負荷から) オブザベーション(←主成分 得点から)	オブザベーション (ほかの手法(←決定木)と 組み合わせて変数も記述で きる)
軸の有無	有 マップの中に方向の概念が ある(ユークリッド空間だから)	無 マップには方向性がないため軸の解釈がされない
次元数	2次元に限定すると考察できるのは2つの主要特性のみ どのぐらい表現できているか→寄与率	マップを広くすれば、オブザベーションの相違を細かく表現→多次元情報を2次元に折りたたんで表現できる

演習1

- kouri.csv(またはretail.csv)を使って 10×10 のSOMを作成してみよう。(前回の復習)

解説1

- マップを作成する

```
setwd("../Dropbox/Share/DL02045/Program/chap4")
```

```
getwd()
```

```
library(som)
```

```
retaildata<-read.csv("kouri.csv",header=T)
```

```
normaldata<-normalize(retaildata[,2:9],byrow=F)
```

```
retailsom<-som(normaldata,xdim=10,ydim=10,topol="rect")
```

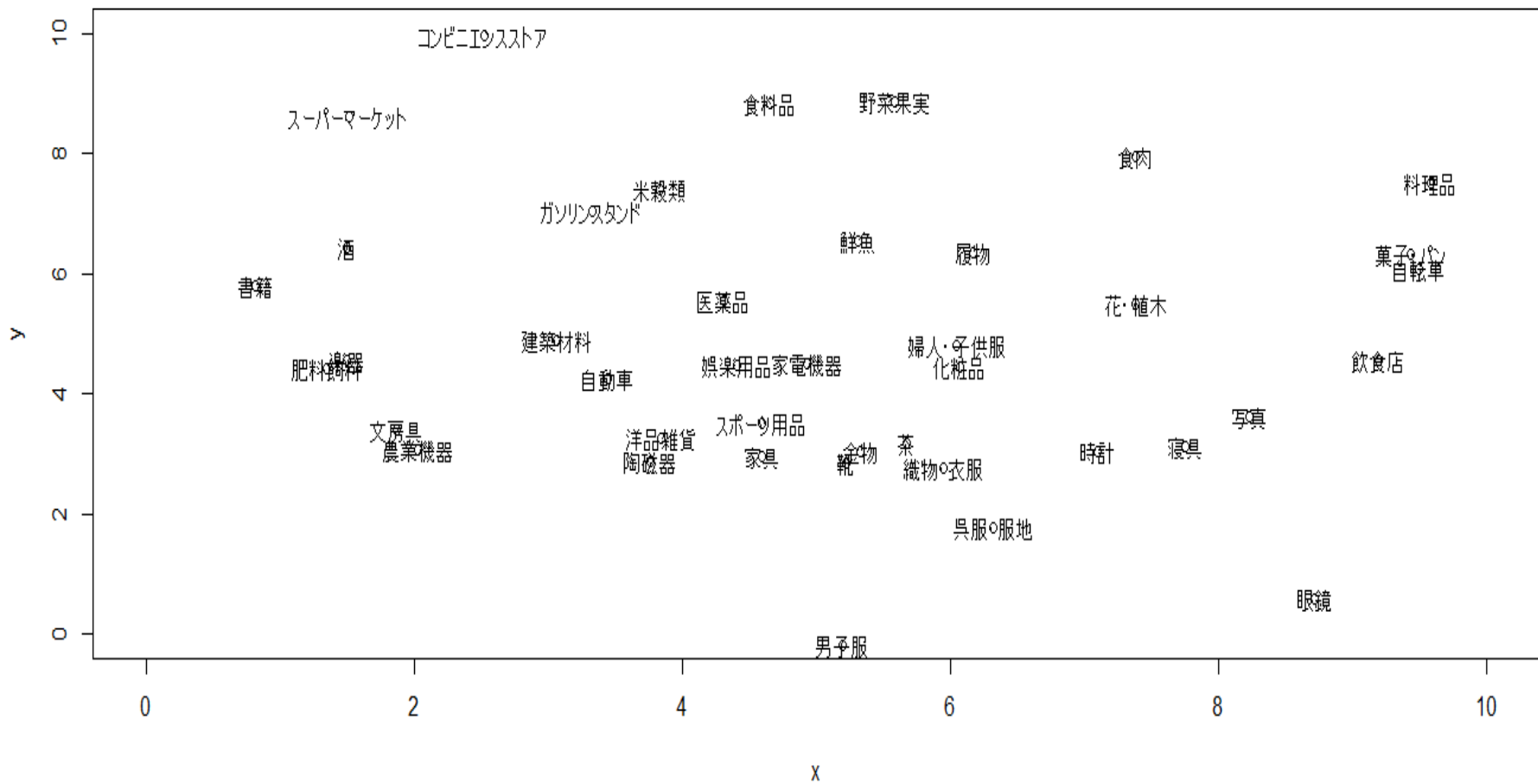
- テキストの表示位置をずらしてマップに座標とテキストを描画する

```
random<-cbind(rnorm(nrow(data),0,0.5),rnorm(nrow(data),0,1.0))
```

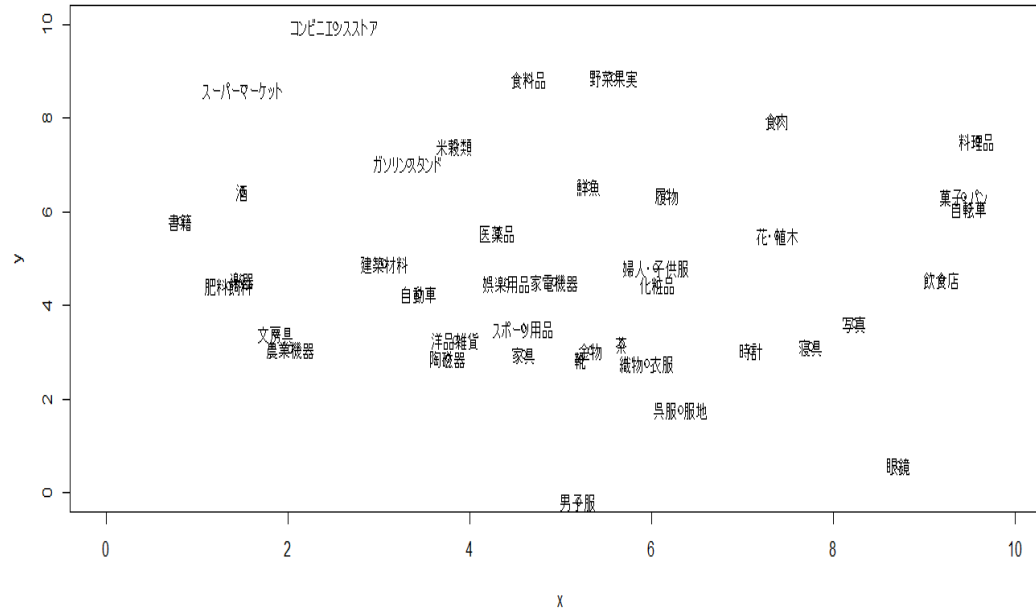
```
map<-retailsom$visual[,1:2]+random+0.5
```

```
plot(map,xlim=c(0,10),ylim=c(0,10))
```

```
text(map[,1],map[,2],data$業種名)
```

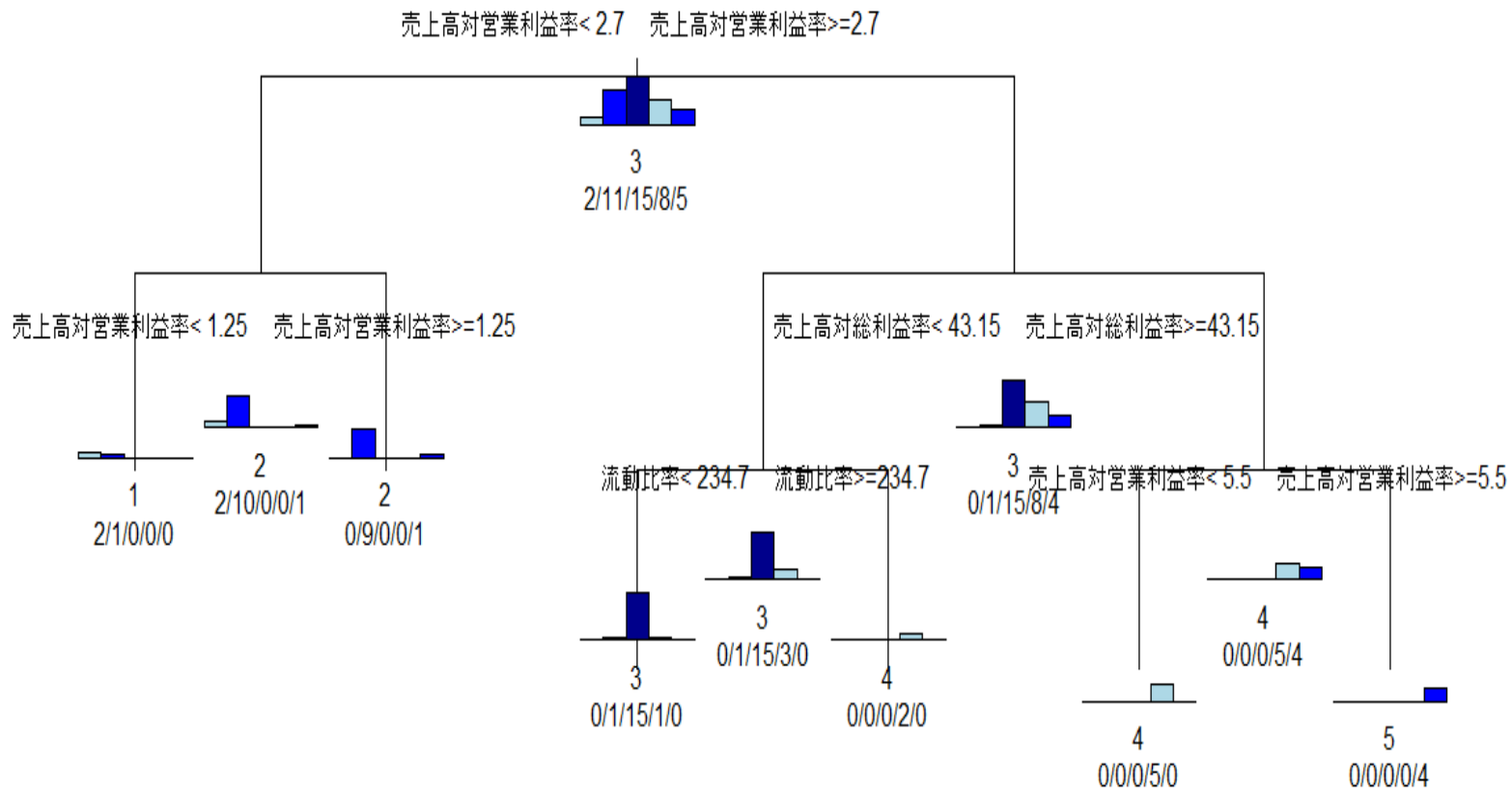


決定木でSOMの表す情報を補う



このマップを見ただけでどこら辺がどういう集まりなのかってパッと見てわかりませんか？

どのような分岐基準でどこのユニットになるのかを求める



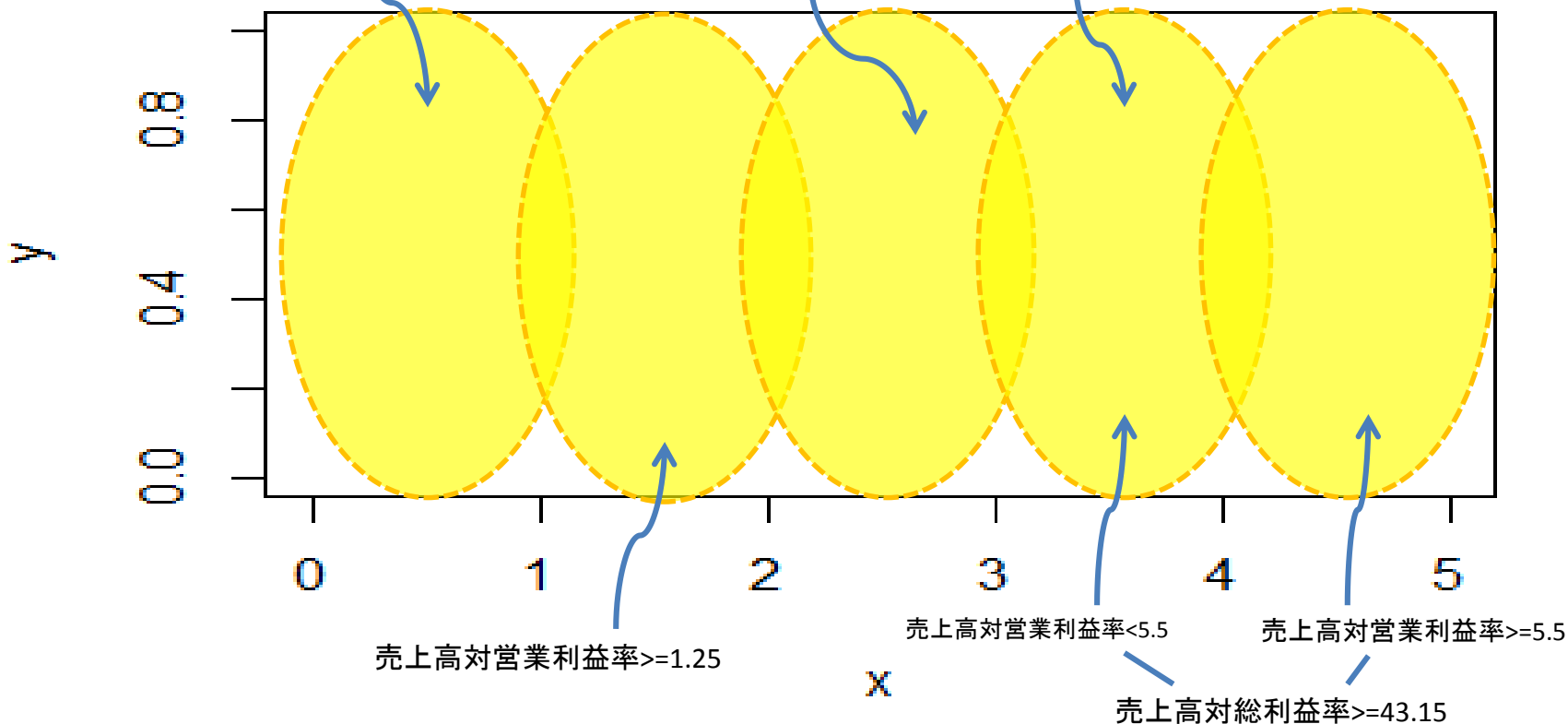
マップ上のどの位置でどういうオブザベーションなのか

売上高対営業利益率 <1.25

売上高対総利益率 <43.15

流動比率 <234.7

流動比率 ≥ 234.7



- つまり、SOMで表現できない情報を決定木を用いて補うことができる。

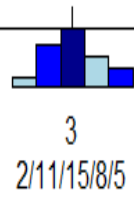
演習2

- SOMは分類の明確な基準が示されない。そこで決定木を利用する。
- SOMで配属されたユニット番号を基準変数として決定木を成長させる。しかし、SOMの100ユニットでは決定木において大変だから、 5×1 にして再分析
- 演習1の内容を決定木で分析してみよう。

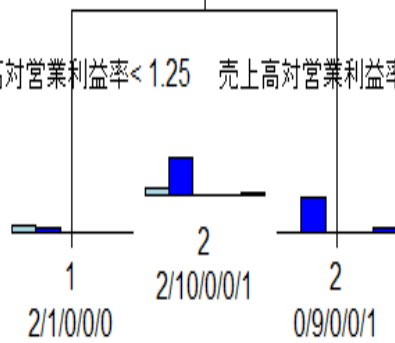
解説2

```
library(mvpart)
retailsom2<-som(normaldata,xdim=5,ydim=1,topol="rect")
retailtreedata<-cbind(retaildata,retailsom2$visual[,1]+1)
colnames(retailtreedata)[10]<-“segment” ←10列目に新しく列
名を与える
tree<-
rpart(segment~.,data=retailtreedata[,2:10],method="class")
print(tree)
plot(tree,uniform=T,branch=1,margin=0.05)
text(tree,all=T,use.n=T)
```

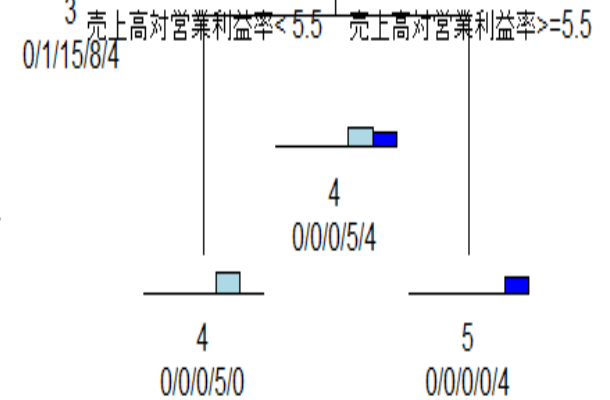
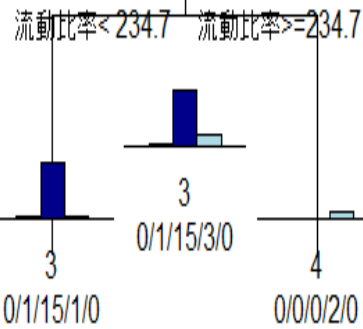
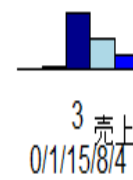
売上高対営業利益率<2.7 売上高対営業利益率>=2.7



売上高対営業利益率<1.25 売上高対営業利益率>=1.25



売上高対総利益率<43.15 売上高対総利益率>=43.15



まとめ・感想

- ・SOMでは各オブザベーションの相関関係が視覚的に見えてくるが、どのような基準で図示されているのか明確には分からない(単純なデータならば考えれば推測できるが)。そこに決定木を用いることでどのような基準でマップを作成しているかが明確にわかるようになる。
- ・他の手法とSOMを比較することでSOMの特徴を深く理解することができた。

宿題

- 何かデータを見つけて、それを用いて自己組織化マップを作成し、決定木を用いて描画結果が何を表してるか考察せよ。

(出力層のユニットの数は任意)

- 扱ったデータファイル(csv など)
- コード、考察
- 結果の画像

をフォルダにまとめて提出