

# 先端メディアゼミナールⅡ

金子曜大

第三回 予測システム

# 第三回でやること

- ゼロサム予測と非ゼロサム予測
- 予測システムをつくる
- 時系列モデルと予測システム
- 演習

ゼロサム予測と非ゼロサム予測

# 予測モデルを扱う

予測モデルを扱う際に、まず初めに重要となるのは「対象の予測は**総和が一定の富を奪い合う行為か否か**」というところである

# ゼロサム予測

総和が一定の富を奪い合う行為の予測

## 特徴

- ・すべての人が儲かることはない
- ・科学的に考えても安定して成功する方法は存在しない

※統計的データ解析が役に立たないと考える人の多くは「予測＝ゼロサム予測」だと思っている

# 非ゼロサム予測(1)

総和が一定の富を奪い合う行為でないものの  
予測(例:商品の売上予測)

## 特徴

- ・上手にデータを集めることで精度が高くなる
- ・ゼロサム予測よりも安定する

# 非ゼロサム予測(2)

予測の分析として他の統計モデルと比べて、精度が高いのが**ニューラルネットを用いた統計モデル**

予測システムをつくる

# 予測変数(1)

予測システムに必ず必要となるのが「**予測変数**」である

予測変数とはその名前の通り**予測するために必要となる変数**のこと

# 予測変数(2)

例:牛乳の販売個数

- 1、当日の気象
- 2、前日までの気象
- 3、曜日
- 4、前日までの客数
- 5、前日までの販売個数
- 6、販売価格

これら1～6の全てが**予測変数**である

# 予測変数(3)

ニューラルネットの長所として、予測変数が基準変数と関係していることが、**主観的にでもわかっている**ならばそれをニューラルネットに組み込むだけで**予測精度が上昇する**という点がある

※**予測変数の入手コスト**と**基準変数の精度の利得**とのバランスを考えてシステムをつくること

# 前処理

教科書p.59 図2.11より

実測値が予測値を大きく上まわっている日がある

この日は大晦日であったり、牛乳以外の商品全体のセールであったり、**予測変数に組み込まれていない事象の影響**である

統計モデルで予測する必要性が小さいものは**担当者が特別に考慮**すればよい

# 時系列モデルと予測モデル

# 時系列モデル作成(1)

## 1、訓練期間

過去のデータをどれくらい利用するか

多: 予測の精度がよくなる

※母集団の変化が起きていたならば変化前のデータは利用しない方がいい

少: 当然予測の精度が低くなる

## 2、訓練データ割合

データの何割を推定用データに利用するか

高: 訓練データより成績が低くなる懸念がある

低: データが少なく推定が不安定になる

# 時系列モデル作成(2)

## 3、使用期間

推定したデータをどれくらいの期間利用するか

長：精度の高いモデルをつくるのが難しい

短：予測モデルのメンテナンスに手間がかかる

## 4、予測期間

何日後の予測をするか

長：精度が悪くなる

短：毎日予測しなければならないため手間がかかる

# 時系列モデル作成(3)

## 5、基準変数

対象のスーパーでは8種類の牛乳を扱っている

A. 予測システムをそれぞれ分けて8つ作る

B. システムは1つにして、それに8つの出力をつくるか

これに関しては実験しないと決着がつかない

「気象」「曜日」などの販売個数に対する影響の仕方が8つとも同じ様な影響ならB、それぞれ独自の影響を受けるならAが望ましい

# 演習

# 演習1(1)

教科書P.59の図2.12のような散布図に回帰直線が加えられた表をつくってみよう

またニューラルネットを用いてグラフに点を打ちこんでみよう。

データは組み込みデータのtreesを用いること

# 演習1(2)

```
plot(Volume~Girth,data=trees)  
result=lm(Volume~Girth,data=trees)  
abline(result)
```

```
a=lm(x~y,data=z)
```

zという名前のデータのxとyを回帰分析してaという変数に格納

```
abline(a)
```

aの回帰直線を描く

# 演習 1 (3)

```
library(nnet)
```

```
n2=nnet(trees$Girth,trees$Volume,linout=T,size  
=30,maxit=2000)
```

```
v.n2=predict(n2,matrix(trees$Girth))
```

```
points(trees$Girth,v.n2,col=4)
```

# まとめと感想

重要となる点は「**予測変数**」と「**前処理**」、そして「**時系列モデル**」

この3つの組み合わせ数は膨大であり、それゆえ、予測モデルを作る際には**モデル構築し続ける過程は必須**となる。

統計的データ解析が役に立たないと考える人もいるが、学んでみるとデータ解析が不可欠であることがわかった

# 宿題

組み込みデータirisのSepal.Lengthをx軸、Petal.Lengthをy軸とした回帰直線を求めよ。  
さらにSepal.Lengthを入力、Petal.Lengthを出力としてニューラルネットを用いて表に描きなさい。