

地理的加重回帰モデルのパラメータ分布について

岡山大学大学院環境学研究科

大下祐樹*

1 地理的加重回帰モデルの結果

地理的重みとして gauss カーネル関数 (バンド幅 1429.33) を用いて地理的加重回帰モデルを構築した。モデルの老人化率の係数の空間分布を図 1 に示す。

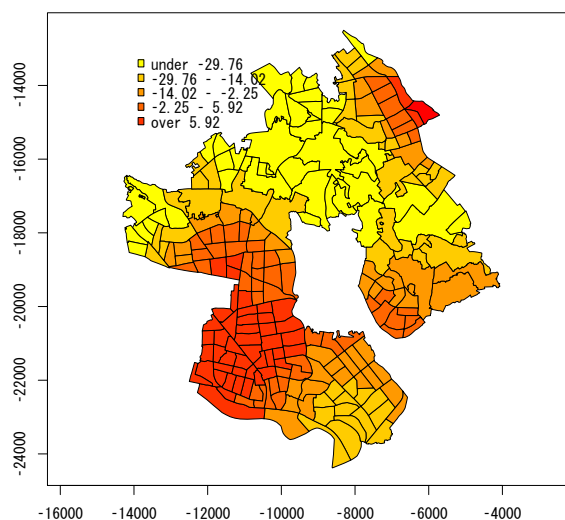


図 1 $\hat{\beta}_{1i}$

図 1 より係数が正の地区が存在する。この正の地区は負となった地区と性格の異なる地区といえる。本研究で用いている変数を表 1 に示す。

R での変数名	sneak	pop65	dh	a.l	a.m	a.h	nc
	空き巣発生率	老人化率	戸建	低層	中層	高層	核家族割合

表 1 の変数を用いてどの程度パラメータが正を判別できるか、判別分析を行った。

2 判別分析

総地区数は 239 である。以降、正の地区は p 、負の地区は n とする。答え合わせ簡便法を用いて、誤判別の個数を k 個として n/k とする。

2.1 すべての変数を用いて

まずは表 1 の変数すべて用いることにする。

```
train <- sample(1:239, 120) #トレーニングデータの地区番号決定
z <- lda(sig ~ pop65 + dh + a.l + a.m + a.h + nc , alldata, prior = c(1,1)/2, subset = train)
pre<-predict(z, alldata[-train, ]) $class #z を用いて予測をする
#答え合わせ
eq<-numeric(239-120)
for(i in 1:119)
{
  if(as.vector(pre)[i]==as.character(alldata$sig[-train][i]))
  {
    eq[i]<-1
  }
}
#正解率
>sum(eq)/119
0.697
```

正解率が約 70 パーセントとなった。したがって、変数すべてを用いて正しく判別できるのは 70 パーセントということである。

2.2 住宅のデータを用いて

表 1 の住宅に関する変数を用いる。

```
train <- sample(1:239, 120) #トレーニングデータの地区番号決定
z <- lda(sig ~ dh + a.l + a.m + a.h , alldata, prior = c(1,1)/2, subset = train)
pre<-predict(z, alldata[-train, ]) $class #z を用いて予測をする
#答え合わせ
eq<-numeric(239-120)
for(i in 1:119)
{
  if(as.vector(pre)[i]==as.character(alldata$sig[-train][i]))
  {
    eq[i]<-1
  }
}
#正解率
sum(eq)/119
0.664
```

正解率が約 66 パーセントとなった。したがって、住宅データを用いて正しく判別できるのは 66 パーセントということである。

2.3 核家族のデータを用いて

表1の住宅に関する変数を用いる。

```
train <- sample(1:239, 120) #トレーニングデータの地区番号決定
z <- lda(sig ~ nc, alldata, prior = c(1,1)/2, subset = train)
pre<-predict(z, alldata[-train, ]) $class #zを用いて予測をする
#答え合わせ
eq<-numeric(239-120)
for(i in 1:119)
{
  if(as.vector(pre)[i]==as.character(alldata$sig[-train][i]))
  {
    eq[i]<-1
  }
}
#正解率
sum(eq)/119
0.672
```

正解率が約67パーセントとなった。したがって、核家族割合のデータを用いて正しく判別できるのは67パーセントということである。

1変数しか用いていないが、高い正解の判別率となった。