

第1回演習課題

統計解析 I (鈴木)

出題: 2018 年 4 月 9 日

期限: 2018 年 5 月 7 日

1. 説明変数と目的変数の N 個の組 $(x_1, y_1), \dots, (x_N, y_N)$ から、 $S := \sum_{i=1}^N (y_i - \beta_0 - \beta_1 x_i)^2$ を最小にする切片 β_0 と傾き β_1 をもとめたい。 $\bar{x} = \frac{1}{N} \sum_{i=1}^N x_i$, $\bar{x^2} = \frac{1}{N} \sum_{i=1}^N x_i^2$, $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$, $\overline{xy} = \frac{1}{N} \sum_{i=1}^N x_i y_i$ を用いて、それらの解 $\hat{\beta}_0, \hat{\beta}_1$ を求めよ。
2. 問題 1 で、各 $i = 1, \dots, N$ に対して、最初に x_i から \bar{x} を、 y_i から \bar{y} をひいて、新しい x_i, y_i に対して計算した \bar{x}, \bar{y} が 0 になるようにしてから $\hat{\beta}_1$ をもとめるにはどうすればよいか。そして、求められた $\hat{\beta}_1$ と、0 にする前のもとの \bar{x}, \bar{y} とを用いて、 $\hat{\beta}_0$ はどのようにかけるか。

以下では、 p 個の説明変数と目的変数の N 個の組 $(x_{1,1}, \dots, x_{1,p}, y_1), \dots, (x_{N,1}, \dots, x_{N,p}, y_N)$ から、切片 β_0 と各変数の傾き $\beta = (\beta_1, \dots, \beta_p)$ を推定したい。各 $x_{i,j}$ から $\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{i,j}$, $j = 1, \dots, p$ をひき、各 y_i から $\bar{y} = \frac{1}{N} \sum_{i=1}^N y_i$ をひき、それぞれの平均を 0 にしてから β を推定する (推定値を $\hat{\beta}$ とする)。最後に $\hat{\beta}_0 = \bar{y} - \sum_{j=1}^p \bar{x}_j \hat{\beta}_j$ とする。また、 $X = (x_{i,j})_{1 \leq i \leq N, 1 \leq j \leq p} \in \mathbb{R}^{N \times p}$, $y = (y_i)_{1 \leq i \leq N} \in \mathbb{R}^N$ とおくものとする。

3. $X^T X$ が正則のときに¹,

$$\|y - X\beta\|^2 := \sum_{i=1}^N (y_i - \beta_1 x_{i,1} - \dots - \beta_p x_{i,p})^2$$

を最小にする β が $\hat{\beta} = (X^T X)^{-1} X^T y$ であたえられることを示せ。

4. 行列 $X \in \mathbb{R}^{n \times p}$ とベクトル $y \in \mathbb{R}$ を入力し、切片の推定値 $\hat{\beta}$ と傾きの推定値 $\hat{\beta} \in \mathbb{R}^p$ の推定値を出力する関数 `linear` を R 言語で構成したい。以下の (1)(2) をうめよ。

```
inner.prod=function(x,y)sum(x*y)
linear = function(X,y){
  n=nrow(X); p=ncol(X);
  X=as.matrix(X); x.bar=array(dim=p); for(j in 1:p)x.bar[j]=mean(X[,j]);
  for(j in 1:p)X[,j]=X[,j]-bar[j]; ## X の中心化
  y=as.vector(y); y.bar=mean(y); y=y-y.bar ## y の中心化
  beta=as.vector(## ここをうめる (1)
  beta.0=## ここをうめる (2)
  return(list(beta=beta,beta.0=beta.0))
}
```

5. λ を非負定数として、

$$\frac{1}{2N} \|y - X\beta\|^2 + \frac{\lambda}{2} \|\beta\|_2^2 := \frac{1}{2N} \sum_{i=1}^N (y_i - \beta_1 x_{i,1} - \dots - \beta_p x_{i,p})^2 + \frac{\lambda}{2} \sum_{j=1}^p \beta_j^2$$

¹ T で転置をあらわす。

を最小にする β をもとめよ (Ridge 回帰)。

6. 問題 4 の関数にさらに、入力としてさらに非負定数 λ を加えて、問題 5 の処理を行う R 言語の関数 `ridge` を求めよ。
7. <https://web.stanford.edu/~hastie/StatLearnSparsity/data.html> の米国犯罪データをダウンロード (その画面で Ctrl-a, Ctrl-c でコピーして、各自 PC のエディタに新規ファイルオープンして、Ctrl-v でペースト) して、`crime.txt` というファイルに格納し、問題 4, 問題 6 で作成したプログラムを実行せよ。

列	説明/目的	変数の意味
1	目的	人口 100 万人あたりの犯罪率
2		(今回は用いない)
3	説明	警察官の年間給与
4	説明	25 歳以上で高校を卒業した人の割合
5	説明	16-19 歳で高校に通っていない人の割合
6	説明	18-24 歳で大学生の割合
7	説明	25 歳以上で 4 年制大学を卒業した人の割合

R を実行する際には、そのファイルが格納されているディレクトリにアクセスできるようにする。

```
> crime=read.table("crime.txt")
> X=crime[,3:7]
> y=crime[,1]
> linear(X,y)
$beta
[1] 10.9806703 -6.0885294  5.4803042  0.3770443  5.5004712
$beta.0
[1] 489.6486
> ridge(X,y)
$beta
[1] 10.9806703 -6.0885294  5.4803042  0.3770443  5.5004712
$beta.0
[1] 489.6486
```

8. $p = 2$ のとき、 $X^T X$ の各成分を、 $v_1 := \sqrt{\frac{1}{N} \sum_{i=1}^N x_{i,1}^2}$, $v_2 := \sqrt{\frac{1}{N} \sum_{i=1}^N x_{i,2}^2}$, $\rho := \frac{\frac{1}{N} \sum_{i=1}^N x_{i,1} x_{i,2}}{v_1 v_2}$ を用いて表示せよ。
9. $X^T X$ の固有値が $\gamma_1, \dots, \gamma_p$ のとき、 $X^T X + N\lambda I$ の固有値が $\gamma_1 + N\lambda, \dots, \gamma_p + N\lambda$ となることを示せ。また、 $X^T X$ に逆行列が存在しない条件を $\gamma_1, \dots, \gamma_p$ を用いてあらわせ。さらに、 $\lambda > 0$ である限り、 $X^T X + N\lambda I$ には逆行列が必ず存在することを示せ。ただし、非負定値²の固有値がすべて非負となることは証明しないでよい。
10. 関数 $f: \mathbb{R}^p \rightarrow \mathbb{R}$ が、任意の $0 < \alpha < 1$ と $u_1, u_2 \in \mathbb{R}^p$ について

$$f(\alpha u_1 + (1 - \alpha) u_2) \leq \alpha f(u_1) + (1 - \alpha) f(u_2)$$

であるとき、関数 f は凸であるという。この定義に基づいて、下記の各関数が凸であることを示せ。

(a) $f(u) = u^2$

² $X^T X$ のように、ある行列 A があって $A^T A$ とかける行列として定義される。

(b) $f(u, v) = u^2 + v^2$

(c) $f(u) = |u|$

11. $f(\beta_1, \dots, \beta_p) := \sum_{i=1}^N (y_i - \sum_{j=1}^p x_{i,j} \beta_j)^2$ が凸であることを示せ。ただし、凸関数の和が凸関数になることは、証明無しで用いてよい。

12. 以下の各関数 $f: \mathbb{R} \rightarrow \mathbb{R}$ と $u_0 \in \mathbb{R}$ について、

$$f(u) \geq f(u_0) + z(u - u_0), \quad u \in \mathbb{R}$$

が成立する $z \in \mathbb{R}$ の集合、および $f(u)$ の $u = u_0$ での微分係数 $f'(u_0)$ を求めよ。今後、この集合を $\partial f(u)$ とかくものとする (劣勾配)。また、要素が 1 個しかない場合、集合ではなく $\{\cdot\}$ をつけず) 要素だけを書く場合がある。

(a) $f(u) = u, u_0 = 0$

(b) $f(u) = u^2 + 2u, u_0 = 1$

また、 $u = u_0$ で微分可能であれば、 $\partial f(u)$ は、その微分係数となることを証明せよ。

13. $f(u) = |u|$ について

$$\partial f(u_0) = \begin{cases} -1, & u_0 < 0 \\ [-1, 1], & u_0 = 0 \\ 1, & u_0 > 0 \end{cases}$$

となることを示せ。

14. λ を非負定数として

$$f(u) = \frac{1}{2N} \sum_{i=1}^N (y_i - ux_{i,1})^2 + \lambda|u|$$

について

$$\partial f(u_0) = \begin{cases} -\frac{1}{N} \sum_{i=1}^N x_{i,1} (y_i - u_0 x_{i,1}) - \lambda, & u_0 < 0 \\ -\frac{1}{N} \sum_{i=1}^N x_{i,1} y_i + \lambda[-1, 1], & u_0 = 0 \\ -\frac{1}{N} \sum_{i=1}^N x_{i,1} (y_i - u_0 x_{i,1}) + \lambda, & u_0 > 0 \end{cases}$$

となることを示せ。また、 $0 \in \partial f(0)$ となる条件を求めよ。ただし、劣勾配の場合でも微分と同様、関数の和の劣勾配がそれぞれの劣勾配の和になること、関数の定数倍の劣勾配がもとの関数の劣勾配の定数倍になることは、証明無しで用いて良い。

15. $p = 1, \frac{1}{N} \sum_{i=1}^N x_{i,1}^2 = 1$ のとき、 λ を非負定数として、

$$\frac{1}{2N} \|y - X\beta\|^2 + \lambda \|\beta\|_1 := \frac{1}{2N} \sum_{i=1}^N (y_i - \beta_1 x_{i,1} - \dots - \beta_p x_{i,p})^2 + \lambda \sum_{j=1}^p |\beta_j|$$

を最小にする $\beta = (\beta_1)$ をもとめよ (Lasso 回帰)。

16. λ を非負定数として、関数

$$S_\lambda(x) := \begin{cases} x - \lambda, & x > \lambda \\ 0, & |x| \leq \lambda \\ x + \lambda, & x < -\lambda \end{cases}$$

が、関数 $(x)_+ = \max\{x, 0\}$,

$$\text{sign}(x) = \begin{cases} -1, & x < 0 \\ 0, & x = 0 \\ 1, & x > 0 \end{cases}$$

を用いて、 $S_\lambda(x) = \text{sign}(x)(|x| - \lambda)_+$ とかけることを示せ。

17. $\lambda > 0, x \in \mathbb{R}$ を入力として、 $S_\lambda(x)$ を出力する関数 `soft.th` を R 言語でかけ。さらに、下記を実行し、関数の動作が正しいことを確認せよ (出力を pdf で提出する)。

```
curve(soft.th(5,x),-10,10)
```

場合分けで関数を定義したり、`max` を用いると、R 言語ではグラフを描いてくれない。`sign`, `abs`, `pmax` を用いよ。

18. 問題 15 の解を関数 $S_\lambda(\cdot)$ を用いてかけ。また、R 言語で記述するとき、下記の空欄をうめよ。

```
linear.lasso.1=function(x,y,lambda){
  x.bar=mean(x); y.bar=mean(y);
  x.scale=scale(x) # 平均化するだけでなく、分散が1になるようにする
  beta=## ここをうめる
  beta.0=y.bar-beta*x.bar
  return(list(beta=beta,beta.0=beta.0))
}
```

19. 下記は、 j 番目の説明変数を固定して、残差

$$y_i - \sum_{k \neq j} x_{i,k} \hat{\beta}_k, \quad i = 1, \dots, N$$

を計算して、係数 $\hat{\beta}_j$ を推定するというのを、 $j = 1, \dots, p$ で繰り返し、さらにそれら全体を繰り返して、 $\hat{\beta}_1, \dots, \hat{\beta}_p$ の値の収束をまつ処理 (Lasso 推定の座標降下法) である。下記の空欄 (残差を計算する箇所) をうめて、 $\lambda = 0$ のときに、関数 `linear` と近い値になることを確認せよ。

```
covar=function(x,y)sum(x*y)/length(x)
linear.lasso=function(X, y, lambda=0, standardize=TRUE){
  X=as.matrix(X)
  p=ncol(X)
  X.bar=array(dim=p); scale=array(dim=p)
  for(j in 1:p){X.bar[j]=mean(X[,j]);X[,j]=X[,j]-X.bar[j];}
  y.bar=mean(y); y=y-y.bar
  beta=array(0, dim=p); beta.old=array(0, dim=p)
  eps=1
  if(standardize==TRUE)for(j in 1:p){scale[j]=sqrt(covar(X[,j],X[,j]));X[,j]=X[,j]/scale[j];}
  while(eps>0.001){
    for(j in 1:p){
      r= ##ここをうめる
      beta[j]=soft.th(lambda,covar(r,X[,j]))/covar(X[,j],X[,j])
    }
    eps=max(abs(beta-beta.old))
    beta.old=beta
  }
  if(standardize==TRUE)for(j in 1:p)beta[j]=beta[j]/scale[j]
  beta.0=y.bar-X.bar[j]*beta[j]
  for(j in 1:p)beta.0=beta.0-X.bar[j]*beta[j]
  return(list(beta=beta, beta.0=beta.0))
}
```

20. $\hat{y} = X\hat{\beta}$ として、 $X^T(y - \hat{y}) = 0$ が成立すること、すなわち $\sum_i x_{i,j}(y_i - \hat{y}_i) = 0, j = 1, \dots, p$ となることを示せ。次に、 $p = 2, c > \sum_{i=1}^n (y_i - \hat{y}_i)^2$ として、 $\beta = (\beta_1, \beta_2)$ に関する方程式 $\|y - X\beta\|^2 = c$ が $\beta = (\hat{\beta}_1, \hat{\beta}_2)$ を中心とする楕円曲線になることを示せ。ただし、必要であれば、以下の等式を用いて良い。

$$\|y - X\beta\|^2 = \sum_{i=1}^N (y_i - \beta_1 x_{i,1} - \beta_2 x_{i,2})^2 = \sum_{i=1}^N \{(y_i - \hat{y}_i) - (\beta_1 - \hat{\beta}_1)x_{i,1} - (\beta_2 - \hat{\beta}_2)x_{i,2}\}^2$$

21. 下記は、犯罪のデータについて、 λ の値を変えながら、各変数の係数がどのように変化するかを表示したものである。plot と lines の行の λ を $\log \lambda$ にかえて、グラフを表示させよ (出力を pdf で提出する)。

```
df=read.table("crime.txt")
x=df[,3:7]
y=df[,1]
p=ncol(x)
lambda.seq=seq(0,2000,10)
plot(log(lambda.seq),coef.seq, ylim=c(-12,12), type="n", col="red") ##この行
for(j in 1:p){
  coef.seq=NULL; for(lambda in lambda.seq)coef.seq=c(coef.seq,ridge(x,y,lambda)$beta[j])
  par(new=TRUE)
  lines(lambda.seq,coef.seq, col=j) ##この行
}
legend("topright",legend=
c("警察官の年間給与","25歳以上で高校を卒業した人の割合",
"16-19歳で高校に通っていない人の割合","18-24歳で大学生の割合",
"25歳以上で4年制大学を卒業した人の割合"), col=1:p, lwd=2, cex=.8)
```

22. ライブラリ glmnet をインストールして、linear.lasso を glmnet におきかえて実行する。glmnet は、 X, y とも行列でないと受け付けない。したがって、

```
X=as.matrix(X); y=as.vector(y)
result=glmnet(X,y)
```

のようにする。result は種々の結果がはいっているが、plot(result) を実行する。

23. 問題 21 の関数 ridge を linear.lasso にかえて実行し、グラフを表示させよ。特に、すべての変数の係数が 0 になるように lambda.seq の範囲を調節せよ (出力を pdf で提出する)。

24. 各 j で、 $\sum_{i=1}^N x_{i,j}^2 = 1$ のとき、すべての変数の係数が 0 になるような λ の最小値が、 $\lambda = \max_{1 \leq j \leq p} \frac{1}{N} \sum_{i=1}^N x_{i,j} y_i$ であたえられることを示せ。