Université de Montpellier Année 2017-2018
HMMA304 - Modèles linéaires généralisés

Examen de contrôle continu
sans document - 11 octobre 2017

Exercice 1 (3 pts)

Expliquer en quoi consiste une regression ridge ?

Exercice 2 (4 pts)

On considère le modèle de régression suivant, pour tout $i = 1, \dots, n$

$$y_i = ax_i + u_i$$

avec $\mathbb{E}(u_i) = 0$, $\mathbb{V}(u_i) = \sigma_i^2$, $\mathbb{C}(u_i, u_j) = 0$ si $i \neq j$, x_i et a des réels.
Donner les expressions des estimateur des moindres carrés ordinaires et généralisés de a.

Exercice 3 (6 pts)

On considère n variables aléatoires indépendantes $y_1, \dots, y_n$ telles que y_i est distribuée suivant une loi inverse gaussienne de paramètres (μ_i, σ^2) où $\mu_i > 0$ et $\sigma^2 > 0$.
Le terme inverse ne doit pas être mal interprété, la loi est inverse dans le sens suivant : la valeur du mouvement brownien à un temps fixé est de loi normale, à l'inverse, le temps en lequel le mouvement brownien avec une dérive positive atteint une valeur fixée est de loi inverse gaussienne. Nous avons

$$f(y_i; \mu_i, \sigma^2) = \frac{1}{\sqrt{2\pi y_i^3}\sigma} \exp\left(-\frac{(y_i - \mu_i)^2}{2(\mu_i\sigma)^2 y_i}\right) \mathbf{1}_{y_i>0}\,.$$

1 (2 pts) Montrer que la loi inverse gausienne appartient à la famille exponentielle avec un paramètre de nuisance.

2 (2 pts) Calculer $\mathbb{E}(y_i)$ et $\mathbb{V}(y_i)$.

Pour tout $i = 1, \dots, n$, on suppose que $\log(\mu_i) = a + bx_i$ où $x_i \in \mathbb{R}$ est supposé connu, a et b sont des paramètres inconnus.

3 (2 pts) Montrer qu'il s'agit d'un modèle linéaire généralisé. La fonction de lien canonique a-t-elle été utilisée ?

Exercice 4 (7 pts)

1 (4 pts) Expliquer en détails les procédures mises en oeuvre par l'intermédiaire du code R suivant.

```
n <- 50
x1 <- rnorm(n)
x2 <- rnorm(n)
x3 <- rnorm(n)
u <- rnorm(n,0,sqrt(2))
d <- 30
z <- matrix(runif(n*d),n)
y <- 1+0.5*x1+2*x2-0.2*x3+u
app <- data.frame(y=y,x1=x1,x2=x2,x3=x3,z=z)
names(app)[-(1:4)] <- paste("z",1:30,sep="")

model1 <- lm(y~1,data=app)
model2 <- lm(y~.,data=app)
model3 <- lm(y~x1+x2+x3,data=app)

library(glmnet)
x <- as.matrix(app[,-1])
model4 <- glmnet(x,y,family="gaussian",nlambda=50,alpha=1)
model5 <- train(x,y,method="glmnet",metric="RMSE",
      trControl=trainControl(method="repeatedcv",
                             number=5,repeats=100),
      tuneGrid=data.frame(alpha=1,lambda=model4$lambda))
```

2 (3 pts) Donner le code R permettant de comparer les taux d'erreurs associés aux modèles considérés dans la question précédente. A quel résultat peut-on s'attendre ?

Correction contrôle continu
HMMA 304
11/10/2017

Exercice 1

Méthode de régularisation en régression utilisée lorsque $n \approx p$ et éventuellement $p > n$

$\hat{\beta}^R$ est la solution de

$$\min_{\beta \in \mathbb{R}^p} \sum_{i=1}^{n} (y_i - x_i \beta)^2 \text{ sous contrainte } \sum \beta_j^2 \leq d$$

Diminuer la variance des estimateurs en introduisant un peu de biais.

Esercizio 2

$$\hat{a}^{OLS} \in \arg\min_{a \in \mathbb{R}} \sum_{i=1}^{n} (y_i - a x_i)^2$$

$$\hat{a}^{OLS} = \frac{\sum_{i=1}^{n} x_i y_i}{\sum_{i=1}^{n} x_i^2}$$

$$\hat{a}^{GLS} \in \arg\min_{a \in \mathbb{R}} \sum_{i=1}^{n} \left[\frac{(y_i - a x_i)^2}{\sigma_i^2} \right]$$

$$\hat{a}^{GLS} = \sum_{i=1}^{n} \frac{x_i y_i}{\sigma_i^2} \Big/ \sum_{i=1}^{n} \frac{x_i^2}{\sigma_i^2}$$

Esercizio 3

1) $$f(y; \mu, \sigma^2) = \exp\left\{ - \frac{y^2 - 2y\mu + \mu^2}{2\mu^2\sigma^2 y} - \frac{1}{2}\log(\sigma^2) \right\}$$

$$\frac{1}{\sqrt{2\pi y^3}} \mathbb{1}\{y > 0\}$$

3

$$f(y;\mu,\sigma^2) = \exp\left\{-\frac{y}{2\mu^2\sigma^2} + \frac{1}{\mu\sigma^2} - \frac{1}{2\sigma^2 y} - \frac{1}{2}\log(\sigma^2)\right\} \frac{1}{\sqrt{2\pi y^3}} \mathbb{1}\{y>0\}$$

$$\theta = -\frac{1}{2\mu^2}, \quad \phi = \sigma^2$$

$$\Rightarrow \mu^2 = \frac{-1}{2\theta} \Rightarrow \mu = \frac{1}{\sqrt{-2\theta}}$$

$$b(\theta) = -\sqrt{-2\theta}$$

$$c(y,\phi) = -\frac{1}{2\phi y} - \frac{1}{2}\log(\phi)$$

la loi inverse gaussienne appartient bien à la famille exponentielle.

2] $$\mathbb{E}(y) = b'(\theta) = \left(-\frac{1}{2}\right)\frac{(-2)}{\sqrt{-2\theta}}$$

$$= \frac{1}{\sqrt{-2\theta}} = \mu$$

4

$V(y) = \phi\, b''(\theta)$

$= \sigma^2 \dfrac{-(-2)}{2(-2\theta)\sqrt{-2\theta}}$

$= \sigma^2 \mu^3$

3) $\log(\mu) = a + b\,x$

$\Rightarrow E(y) = \mu = e^{a+bx}$

$\Rightarrow E(y) = g(a+bx)$

Il s'agit bien d'un modèle linéaire généralisé

$g(\mu) = e^{\mu} \neq b'(\mu)$

Ce n'est pas le lien canonique qui a été utilisé.