Modern Regression: Examination 2019

25 June 2019

Instructions: The time allotted for the examination is 180 minutes. You may answer in either English or French. No written material may be brought into the examination, but a simple calculator may be used. Full marks may be obtained with complete answers to four questions. The final mark will be based on the best four solutions.

Notation: $a_+ = \max(a, 0)$ for $a \in \mathbb{R}$; $A_{r\times s}$ means that A is an $r \times s$ matrix; $X \sim \mathcal{N}_p(\mu, \Omega)$ means that X has a p-dimensional multivariate normal distribution with mean vector $\mu_{p\times 1}$ and variance matrix $\Omega_{p\times p}$; and $X_{p\times 1} \sim (\mu, \Omega)$ means that $\mathrm{E}(X) = \mu_{p\times 1}$ and $\mathrm{var}(X) = \Omega_{p\times p}$.

First name:

Last name:

SCIPER number:

Exercise	Points	Indicative marks
1		/10 points
2		/10 points
3		/10 points
4		/10 points
5		/10 points
Total:		/40 points

1. Independent random variables $Y_1, \ldots, Y_n$ have probability density functions

$$f(y_j; \beta) = \frac{1}{\pi \left\{ 1 + \left(y_j - x_j^{\mathrm{T}} \beta \right)^2 \right\}}, \quad y_j \in \mathbb{R}, \quad j = 1, \ldots, n,$$

where β is a $p \times 1$ vector of unknown real-valued parameters and $x_1, \ldots, x_n$ are $p \times 1$ vectors of explanatory variables.

(a) Derive an iterative weighted least squares algorithm to obtain the maximum likelihood estimator of β.

(b) How would you modify your approach if $y_j - x_j^{\mathrm{T}}\beta$ was replaced by $(y_j - x_j^{\mathrm{T}}\beta)/\sigma$?

(c) Let $a(u) = \mathrm{d}^2 \log(1 + u^2)/\mathrm{d}u^2$. Show that if the random variable U has probability density function $\{\pi(1 + u^2)\}^{-1}$ for $u \in \mathbb{R}$, then

$$\Pr\{a(U) > 0\} = 1/2,$$

explain what implications this has for your algorithm, and outline how you could overcome them.

2. (a) In what senses does a *generalized linear model* extend the range of application of the linear model? Give two examples of generalized linear models.

(b) Show that the chi-squared density with known degrees of freedom ν,

$$\frac{y^{\nu/2-1}}{2^{\nu/2}\sigma^{\nu}\Gamma(\nu/2)} \exp\left(-\frac{y}{2\sigma^2}\right), \quad y > 0, \sigma > 0, \nu = 1, 2, \ldots,$$

can be written in generalized linear model form

$$f(y; \theta, \phi) = \exp\left\{ \frac{y\theta - b(\theta)}{\phi} + c(y; \phi) \right\},$$

where θ and ϕ are functions, to be found, of ν and σ^2. Compute the mean and variance of y.

(c) The yield of an industrial process was measured r_i times independently at m different temperatures t_i. The resulting yields Z_{ij}, $(j = 1, \ldots, r_i, i = 1, \ldots, m)$, may be assumed to be independent and normally distributed with both means ζ_i and variances τ_i dependent on t_i. Explain how the sums of squares $Y_i = \sum_{j=1}^{r_i} (Z_{ij} - \bar{Z}_i)^2$, where $\bar{Z}_i = r_i^{-1} \sum_{j=1}^{r_i} Z_{ij}$, may be used to assess the dependence of variance on temperature in a suitable generalized linear model. Briefly discuss the advantages and disadvantages of the canonical link function of your model.

3. The output below is from the analysis of data on the presence of calcium oxalate crystals in 77 samples of urine. The binary response r is an indicator of the presence of such crystals, and there are six explanatory variables: specific gravity (`grav`), i.e., the density of urine relative to water; pH (`ph`); osmolarity (`osmo`, mOsm); conductivity (`cond`, mMho); urea concentration (`urea`, millimoles per litre); and calcium concentration (`calc`, millimoles per litre).

   (a) What model has been fitted to the data? How is the response variable related to the explanatory variables?

   (b) How do you interpret the analysis of deviance table? What actions does it suggest to you?

   (c) Compare the deviance reduction due to `grav` and the corresponding estimated regression coefficient. What does this suggest to you?

   (d) Give a 95% confidence interval for the parameter corresponding to `calc`.

   (e) How does a fitted value change if `ph` is decreased by 0.1?

   (f) What can you say about the fit of the model, based on this output?

```
> anova(urine.glm)
Analysis of Deviance Table

Model: binomial, link: logit

Terms added sequentially (first to last)

                       Df Deviance Resid. Df Resid. Dev
NULL                                      76    105.168
grav                    1  14.9327        75     90.235
ph                      1   0.0723        74     90.163
osmo                    1   9.5573        73     80.606
cond                    1   0.0106        72     80.595
urea                    1   1.3343        71     79.261
calc                    1  21.7007        70     57.560

> summary(urine.glm)
Call:
glm(formula = r ~ grav + ph + osmo + cond + urea + calc, family = binomial, data = urine)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.6215  -0.5967  -0.2849   0.3176   2.7445

Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
(Intercept)             0.60609    3.79582   0.160  0.87314
grav                    3.55944    2.22110   1.603  0.10903
ph                     -0.49570    0.56976  -0.870  0.38429
osmo                    0.01681    0.01782   0.944  0.34536
cond                   -0.43282    0.25123  -1.723  0.08493 .
urea                   -0.03201    0.01612  -1.986  0.04703 *
calc                    0.78369    0.24216   3.236  0.00121 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 105.17  on 76  degrees of freedom
Residual deviance:  57.56  on 70  degrees of freedom
AIC: 71.56
```

4. Consider a regression model of the form

$$y_{n\times 1} = X_{n\times p}\beta_{p\times 1} + Z_{n\times q}b_{q\times 1} + \varepsilon_{n\times 1}, \quad \varepsilon \sim \mathcal{N}_n(0,\Omega) \perp\!\!\!\perp b \sim \mathcal{N}_q(0,\Omega_b).$$

(a) Data y, X and Z are available, and it is desired to predict b by choosing $\tilde{b}(y)$ to minimise

$$\mathrm{E}\left[\{\tilde{b}(y)-b\}^{\mathrm{T}}\{\tilde{b}(y)-b\}\right].$$

Show that $\tilde{b}(y) = \mathrm{E}(b \mid y)$.

(b) Find the joint distribution of y and b, and hence show that

$$\begin{aligned}\mathrm{E}(b \mid y) &= \left(Z^{\mathrm{T}}\Omega^{-1}Z + \Omega_b^{-1}\right)^{-1} Z^{\mathrm{T}}\Omega^{-1}(y - X\beta),\\ \mathrm{var}(b \mid y) &= \left(Z^{\mathrm{T}}\Omega^{-1}Z + \Omega_b^{-1}\right)^{-1}.\end{aligned}$$

How would these formulae be used in practice?

(c) Discuss what happens when $\sigma_b^2/\sigma^2 \gg 1$ and $\sigma_b^2/\sigma^2 \ll 1$ in the special case $\Omega = \sigma^2 I_n$ and $\Omega_b = \sigma_b^2 I_q$.

Recall (i) that $(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1}+DA^{-1}B)^{-1}DA^{-1}$ for compatible matrices A, B, C, D and if the necessary inverses exist, and (ii) that if

$$\begin{pmatrix} Y_{\mathcal{A}} \\ Y_{\mathcal{B}} \end{pmatrix} \sim \mathcal{N}\left\{\begin{pmatrix} \mu_{\mathcal{A}} \\ \mu_{\mathcal{B}} \end{pmatrix}, \begin{pmatrix} \Omega_{\mathcal{A}} & \Omega_{\mathcal{A},\mathcal{B}} \\ \Omega_{\mathcal{B},\mathcal{A}} & \Omega_{\mathcal{B}} \end{pmatrix}\right\},$$

then the distribution of $Y_{\mathcal{A}}$ conditional on $Y_{\mathcal{B}} = y_{\mathcal{B}}$ is $\mathcal{N}\left\{\mu_{\mathcal{A}} + \Omega_{\mathcal{A},\mathcal{B}}\Omega_{\mathcal{B}}^{-1}(y_{\mathcal{B}} - \mu_{\mathcal{B}}), \Omega_{\mathcal{A}} - \Omega_{\mathcal{A},\mathcal{B}}\Omega_{\mathcal{B}}^{-1}\Omega_{\mathcal{B},\mathcal{A}}\right\}$.

5. (a) Under what circumstances would you consider the use of a model of the form

$$y_j = \mu(x_j) + \varepsilon_j, \quad j = 1,\ldots,n,$$

where the ε_j are independent random variables and $\mu(x)$ is a suitably smooth function of the scalar x? Describe and compare three methods for choosing a suitable degree of smoothness of μ.

(b) The result of minimising the penalized sum of squares

$$\sum_{j=1}^{n}\{y_j - \mu(x_j)\}^2 + \lambda\int \mu''(x)^2\,\mathrm{d}x$$

is a function $\hat{\mu}(x)$ that yields the $n\times 1$ vector of fitted values $\hat{\mu} = Sy$, with $\hat{\mu}(x_j)$ the jth element of $\hat{\mu}$. Let $\hat{\mu}^{-j}(x)$ and $\hat{\mu}^{-j}$ denote the corresponding function and $n\times 1$ vector of fitted values when the jth pair (x_j, y_j) is not included in the fit. Define an $n\times 1$ vector y^* by setting $y_i^* = y_i$ for $i \neq j$ and $y_j^* = \hat{\mu}^{-j}(x_j)$. By considering the inequality

$$\sum_{i=1}^{n}\{y_i^* - \mu(x_i)\}^2 + \lambda\int \mu''(x)^2\,\mathrm{d}x \geq \sum_{i\neq j}^{n}\{y_i^* - \mu(x_i)\}^2 + \lambda\int \mu''(x)^2\,\mathrm{d}x,$$

show that $\hat{\mu}^{-j} = Sy^*$.

(c) Deduce that

$$y_j - \hat{\mu}^{-j}(x_j) = \frac{y_j - \hat{\mu}(x_j)}{1 - S_{jj}}, \quad j = 1,\ldots,n.$$

Explain how this expression is useful in choosing λ.