STA 223 Homework 5

Bohao Zou (bhzou@ucdavis.edu)

February 24, 2021

1. Problem set 3, 4

Because $X^{'} = W^{\frac{1}{2}}X$ and $Z^{'} = W^{\frac{1}{2}}Z$, we can know that $X = W^{-\frac{1}{2}}X^{'}$ and $Z = W^{-\frac{1}{2}}Z^{'}$. The weighted least squares updating step in IWLS is:

$$\hat{\beta}_{l+1} = argmin_{\beta} \quad (Z_l - X\beta)^T W_l (Z_l - X\beta)$$

where l represents the iteration time. By using $X = W^{-\frac{1}{2}}X^{'}$ and $Z = W^{-\frac{1}{2}}Z^{'}$, we can get another formula.

$$\begin{aligned}
\hat{\beta}_{l+1} &= argmin_{\beta} \quad (W^{-\frac{1}{2}}Z^{'} - W^{-\frac{1}{2}}X^{'}\beta)^T W (W^{-\frac{1}{2}}Z^{'} - W^{-\frac{1}{2}}X^{'}\beta) \\
&= (Z^{'} - X^{'}\beta)^T W^{-\frac{1}{2}} W W^{-\frac{1}{2}} (Z^{'} - X^{'}\beta) \\
&= (Z^{'} - X^{'}\beta)^T (Z^{'} - X^{'}\beta)
\end{aligned}$$

This is the form of unweighted least squares.

2. Problem set 3, 5

Observing that $A^T = A$ and $A^2 = A$ shows that the projection is an orthogonal projection.

$$\begin{aligned}
H &= W^{\frac{1}{2}}X(X^TWX)^{-1}X^TW^{\frac{1}{2}} \\
H^T &= (W^{\frac{1}{2}})^TX[(X^TWX)^{-1}]^TX^T(W^{\frac{1}{2}})^T \\
&= W^{\frac{1}{2}}X[(X^TWX)^T]^{-1}X^TW^{\frac{1}{2}} \\
&= W^{\frac{1}{2}}X(X^TWX)^{-1}X^TW^{\frac{1}{2}} = H \\
H^2 &= W^{\frac{1}{2}}X(X^TWX)^{-1}X^TW^{\frac{1}{2}}W^{\frac{1}{2}}X(X^TWX)^{-1}X^TW^{\frac{1}{2}} \\
&= W^{\frac{1}{2}}X(X^TWX)^{-1}X^TWX(X^TWX)^{-1}X^TW^{\frac{1}{2}} \\
&= W^{\frac{1}{2}}X(X^TWX)^{-1}X^TW^{\frac{1}{2}} = H
\end{aligned}$$

From the above prove, we can know the hat matrix $H = H^T$ and $H^2 = H$. So, the hat matrix indeed is a projection matrix for an orthogonal projection.

3. Problem set 3, 6

The weight matrix is W $= diag(g^{'}(\mu_1)^2V(\mu_1)\phi, \ldots, g^{'}(\mu_n)^2V(\mu_n)\phi)^{-1}$, where $diag$ means the diagnoal matrix, function g means the link function and V means the variance function and ϕ is a coefficient of PMF or PDF. The information matrix $I = -E[H] = -X^TWX$, where X is the design matrix and the W is the weight matrix.

Logistic model

The link function g, variance function V and coefficient ϕ of logistic model are $g'(\mu) = \frac{1}{(1-\mu)\mu}$, $V(\mu) = \mu(1-\mu)$, $\phi = \frac{1}{n}$. So, the i-th diagnoal element of weight matrix w_{ii} is

$$w_{ii} = \frac{1}{n\mu_i(1-\mu_i)}$$

Poisson model

The link function g, variance function V and coefficient ϕ of logistic model are $g'(\mu) = \frac{1}{\mu}$, $V(\mu) = \mu$, $\phi = 1$. So, the i-th diagnoal element of weight matrix w_{ii} is

$$w_{ii} = \frac{1}{\mu_i}$$

By using those w_{ii}, we can get the weight matrix and corresponding information matrix.

4. Problem set 4, 9

In this question, I used binomial regression model and the predictors are all included into the regression model to build GAM. The smoothing function which I uesd is $s()$ function in R with $df = 5$. The table of Anova for Parametric Effects is showed below:

term	df	sumsq	meansq	statistic	p.value
s(fixed.acidity, 5)	1	7.691632	7.691632	8.167442	0.004322
s(volatile.acidity, 5)	1	54.005962	54.005962	57.346811	0
s(citric.acid, 5)	1	2.635662	2.635662	2.798706	0.094543
s(residual.sugar, 5)	1	0.207452	0.207452	0.220285	0.638889
s(chlorides, 5)	1	1.34998	1.34998	1.433491	0.23138
s(free.sulfur.dioxide, 5)	1	5.533537	5.533537	5.875846	0.015464
s(total.sulfur.dioxide, 5)	1	33.003193	33.003193	35.044795	0
s(density, 5)	1	67.226691	67.226691	71.385384	0
s(pH, 5)	1	4.824601	4.824601	5.123055	0.023748
s(sulphates, 5)	1	85.478008	85.478008	90.765741	0
s(alcohol, 5)	1	61.570285	61.570285	65.379069	0

The table of Anova for non-Parametric Effects is showed below:

s(fixed.acidity, 5)	4	14.488	0.0058895
s(volatile.acidity, 5)	4	5.913	0.2057123
s(citric.acid, 5)	4	4.929	0.2946134
s(residual.sugar, 5)	4	9.354	0.0528246
s(chlorides, 5)	4	16.298	0.0026437
s(free.sulfur.dioxide,	4	10.427	0.0338207
s(total.sulfur.dioxide	4	18.651	0.0009202
s(density, 5)	4	11.679	0.0199083
s(pH, 5)	4	6.273	0.1796601
s(sulphates, 5)	4	40.6	3.25E-08
s(alcohol, 5)	4	24.613	6.02E-05

From those two tables we can know under the significance level $\alpha = 0.05$, the additive functions for the predictors that are included after predictor selection are "fixed.acidity", "chlorides", "free.sulfur.dioxide", "total.sulfur.dioxide", "density", "sulphates" and "alcohol". If one of smoothing function $f_j(x_j)$ is not significant, this may indicate that the predictor x_j is linear with response variable. So, we can remove the smoothing function f_j for this predictor x_j.

5. Problem set 4, 12

(a)

The pmf of $Poisson(\mu)$ by $f(s;\mu) = \frac{\mu^s}{s!}e^{-\mu}$ for s = 0, 1, 2.... The components of $ten-inflated$ Poisson model will be used here. The linear predictors $\eta = X\beta$ and $\tilde{\eta} = X\gamma$. The link functions are $\eta = log(\mu)$ and $\tilde{\eta} = logit(\alpha)$. The random components $P(s_i = y|X_i) = \alpha_i \mathbf{1}_{y=10} + (1-\alpha_i)f(y;\mu_i)$

(b)

Fitting the model to data can max the formula

$$max_{\beta,\gamma} l(\beta,\gamma;s,X) = max_{\beta,\gamma}\sum_{i=1}^{n} log[expit(X_i^T\gamma)\mathbf{1}_{s_i=10} + (1-expit(X_i^T\gamma))f(s_i;exp(X_i^T\beta))]$$

Then, the max can be obtained by Newton-Raphson method.

(c)

The estimate for fraction of recording affected at covariated level X_i is given by

$$\frac{\hat{\alpha_i}}{\hat{\alpha_i} + (1-\hat{\alpha_i})f(10;\hat{\mu_i})}$$

where $\hat{\mu_i} = exp(X_i^T\hat{\beta})$ and $\hat{\alpha_i} = expit(X_i^T\hat{\gamma})$

6. Problem set 4, 13

The estimated coefficients of poisson regression are showed in the table:

term	estimate	std.error	statistic	p.value
(Intercept)	-76.107002	6.019832	-12.642711	0
years	0.040588	0.003075	13.200574	0
sunspotnumber	0.000574	6e-04	0.957338	0.338397

From this table we can know the intercept and years are all significant under 0.05 level. The sun spot number is not significant.

By using the random X bootstrap for this data with Poisson regression model and set $B = 2000$. The 95% confidence interval for years is $[0.03575623, 0.04575644]$. The 95% confidence interval for sun spot number is $[-0.0003674874, 0.0014595242]$. From those C.I. we can know we have the same conclusion with previous model.

7. Problem set 4, 14

The estimated coefficients and 95% confidence interval for each coefficients are showed in the table:

term	estimate	std.error	statistic	p.value	lwb 2.50%	upb 97.50%
(Intercept)	-8.404696	0.716636	-11.727987	0	-9.860319374	-7.048106262
V1	0.123182	0.032078	3.84014	0.000123	0.060918463	0.186855824
V2	0.035164	0.003709	9.481393	0	0.028092756	0.042650074
V3	-0.013296	0.005234	-2.540416	0.011072	-0.023682464	-0.003103975
V4	0.000619	0.006899	0.089713	0.928515	-0.01284946	0.014211576
V5	-0.001192	0.000901	-1.322309	0.186065	-0.002966884	0.000582143
V6	0.089701	0.015088	5.945334	0	0.060849478	0.12006085
V7	0.94518	0.299147	3.159578	0.00158	0.365370025	1.538656174
V8	0.014869	0.009335	1.592858	0.111192	-0.003503266	0.033186571

The 95% confidence intervals from the random X bootstrap for this data are showed below:

	lwb 2.50%	upb 97.50%
(Intercept)	-10.25076941	-7.162474103
V1	0.055358978	0.200224557
V2	0.028063423	0.044897258
V3	-0.023905018	-0.003514257
V4	-0.013295522	0.014438231
V5	-0.003276744	0.000683906
V6	0.062446582	0.124479162
V7	0.285433087	1.68546461
V8	-0.004754981	0.036147999

Compare with those confidence interval which from MLE and X bootstrap, we can know that the C.I.s are all roughly same.

8. Problem set 5, 3

The constant-coefficient-of-variation model has property that $c = \frac{Var(y)}{(E[y])^2}$, where c is a constant. $Var(y) = \sigma^2 V(\mu)$ and $\eta = g(\mu)$ and $E[y] = \mu$. From those we can know $\frac{Var(y)}{(E[y])^2} = \frac{\sigma^2 V(\mu)}{\mu^2} = c$, $\sigma^2 V(\mu) = c\mu^2$. By this property we can know the Quasi-score is

$$U = u(\mu; y) = \frac{y-\mu}{\sigma^2 V(\mu)} = \frac{y-\mu}{c\mu^2}$$

By using this, we can define Quasi-Likelihood(QL) for one observation:

$$\begin{aligned} Q(\mu, y) &= \int_y^\mu \frac{y-t}{ct^2} dt \\ &= \frac{1}{c}\left(\int_y^\mu \frac{y}{t^2} dt - \int_y^\mu \frac{1}{t} dt\right) \\ &= \frac{1 - \frac{y}{\mu} - log(\frac{\mu}{y})}{c} \\ &= \frac{1}{c}\left(1 - \frac{y}{\mu} + log(\frac{y}{\mu})\right) \end{aligned}$$

By using this QL for one observation, we can get QL for all observation

$$Q_{All}(\mu, y) = \sum_{i=1}^{n} Q_i(\mu_i, u_i) = \frac{1}{c}(n - \sum_{i=1}^{n} \frac{y_i}{\mu_i} + \sum_{i=1}^{n} log(\frac{y_i}{\mu_i}))$$

9. Problem set 5, 6

We can know $U_j(\beta) = \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta_j} \frac{y_i - \mu_i}{\sigma^2 V(\mu_i)}$ for $1 \leq j \leq p$. For the j-th row and r-th coloum element of information matrix $i_\beta = -E[\frac{\partial U}{\partial \beta}]$ we can have $I_{jr} = -E[\frac{\partial U_j(\beta)}{\partial \beta_r}]$. We can rewrite $U_j(\beta)$ to this form:

$$\begin{aligned} U_j(\beta) &= \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta_j} \frac{y_i - \mu_i}{\sigma^2 V(\mu_i)} \\ &= \sum_{i=1}^{n} f(\mu_i)(y_i - \mu_i) \\ f(\mu_i) &= \frac{\partial \mu_i}{\partial \beta_j} \frac{1}{\sigma^2 V(\mu_i)} \end{aligned}$$

For the $\frac{\partial U_j(\beta)}{\partial \beta_r}$ we can get

$$\begin{aligned} \frac{\partial U_j(\beta)}{\partial \beta_r} &= \sum_{i=1}^{n} [f'(\mu_i)(y_i - \mu_i) + f(\mu_i)\frac{\partial (y_i - \mu_i)}{\partial \beta_r}] \\ I_{jr} &= -E[\frac{\partial U_j(\beta)}{\partial \beta_r}] = -\sum_{i=1}^{n} f(\mu_i)\frac{\partial (y_i - \mu_i)}{\partial \beta_r} \\ &= -\sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta_j} \frac{1}{\sigma^2 V(\mu_i)} \frac{\partial (y_i - \mu_i)}{\partial \beta_r} \\ &= \frac{1}{\sigma^2} \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta_j} \frac{1}{V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_r} \end{aligned}$$

From this we can know the j-th row and r-th column element of information matrix i_β is $I_{jr} = \sum_{i=1}^{n} \frac{1}{\sigma^2} \frac{\partial \mu_i}{\partial \beta_j} \frac{1}{V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_r}$. So, for the information matrix i_β we can know $i_\beta = D^T V^{-1} D / \sigma^2$.
Because we know $\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j}$. We can further improve the formula of I_{jr}:

$$\begin{aligned} I_{jr} &= \sum_{i=1}^{n} \frac{\partial \mu_i}{\partial \beta_j} \frac{1}{\sigma^2} \frac{1}{V(\mu_i)} \frac{\partial \mu_i}{\partial \beta_r} \\ &= \sum_{i=1}^{n} x_{ij} \frac{1}{\sigma^2} \frac{1}{V(\mu_i)} (\frac{\partial \mu_i}{\partial \eta_i})^2 x_{ir} \end{aligned}$$

From this formula we can know the information matrix $i_\beta = X^T W X$, where W is a diagnoal matrix which

$$w_{ii} = \frac{1}{g'(\mu_i) \phi V(\mu_i)} = \frac{1}{\phi} \frac{\partial \theta_i}{\partial \mu_i} (\frac{\partial \mu_i}{\partial \eta_i})^2$$