STA 223 Homework 6

Bohao Zou (bhzou@ucdavis.edu)

March 3, 2021

1. Problem 10

$$
\begin{aligned}
E(Y|X_j) &= E(E(Y|X_1, ..., X_p)|X_j) \\
&= E(\alpha + \sum_k g_k(x_k)|X_j) \\
&= \alpha + E(\sum_{k \neq j} g_k(x_k)) + E(g_j(x_j)|X_j) \quad \# \quad X_i \quad are \quad independent \\
&= \alpha + 0 + g_j(x_j) \\
&= \alpha + g_j(x_j)
\end{aligned}
$$

We can also know $E[Y] = E(E(Y|X)) = E(\alpha + \sum_j g_j(x_j)) = \alpha$. The simple method is that we can first estimate $\hat{\alpha} = E(Y)$ and then for each X_j, we can use one dimensional smoothing and make $E(g_j(x_j)) = 0$.

For situations that if we know the predictors are all independent with each other, we can use this method to fit model without backfitting.

2. Problem 11

(a)

The precise Bayes formula is

$$P(Y = k|X = x) = \frac{P(X = x|Y = k)P(Y = k)}{\sum_j P(X = x|Y = j)P(Y = j)}$$

Because we know f_0, f_1 are the pdf (or the pmf) of covariates X for the subjects belonging to group $G0$ or $G1$. Assuming marginal probabilities $\pi_1 = P(Y = 1), \pi_0 = P(Y = 0)$, we can get the conditional probability:

$$P(Y = k|X = x) = \frac{f_k(x)\pi_k}{f_0(x)\pi_0 + f_1(x)\pi_1}$$

(b)

Classify into $G1$ if

$$k = \frac{P(Y = 1|X = x)}{P(Y = 0|X = x)} = \frac{f_1(x)\pi_1}{f_0(x)\pi_0} > 1$$

If we give the log transformation to k, then work with the equivalent version:

$$
\begin{aligned}
&k > 1 \\
&log(k) > log(1) = 0 \\
&log(\frac{P(Y=1|X=x)}{P(Y=0|X=x)}) > 0
\end{aligned}
$$

(c)

If $\Sigma_k = \Sigma$ for $k = 0, 1$, then the classification rule becomes:

$$
\begin{aligned}
&log(\frac{P(Y=1|X=x)}{P(Y=0|X=x)}) = log(\frac{f_1(x)\pi_1}{f_0(x)\pi_0}) \\
&= log(\frac{\pi_1(2\pi^{-p/2})det(\Sigma)^{-1/2}exp(-1/2(x-\mu_1)^T\Sigma^{-1}(x-\mu_1))}{\pi_0(2\pi^{-p/2})det(\Sigma)^{-1/2}exp(-1/2(x-\mu_0)^T\Sigma^{-1}(x-\mu_0))}) \\
&= log(\frac{\pi_1}{\pi_0}) + log(exp(-\frac{1}{2}(x-\mu_1)^T\Sigma^{-1}(x-\mu_1) + \frac{1}{2}(x-\mu_0)^T\Sigma^{-1}(x-\mu_0))) \\
&= log(\frac{\pi_1}{\pi_0}) + \frac{1}{2}(\mu_1-\mu_0)\Sigma^{-1}(2x-(\mu_1+\mu_0)) \\
&= log(\frac{\pi_1}{\pi_0}) + (x-\frac{1}{2}(\mu_1+\mu_0))^T\Sigma^{-1}(\mu_1-\mu_0)
\end{aligned}
$$

(d)

$$
log(\frac{P(Y=1|X=x)}{P(Y=0|X=x)}) = log(\frac{P(Y=1|X=x)}{1-P(Y=1|X=x)})
$$

We can set $P(Y=1|X=x) = \mu$, then

$$
log(\frac{P(Y=1|X=x)}{1-P(Y=1|X=x)}) = log(\frac{\mu}{1-\mu}) = logit(\mu)
$$

Because $E(Y|X=x) = P(Y=1|X=x) * 1 + 0 * P(Y=0|X=x)$, we can get

$$
logit(\mu) = logit(P(Y=1|X=x)) = logit(E(Y|X=x))
$$

(e)

Logistic classifier is also a Bayes classifier.

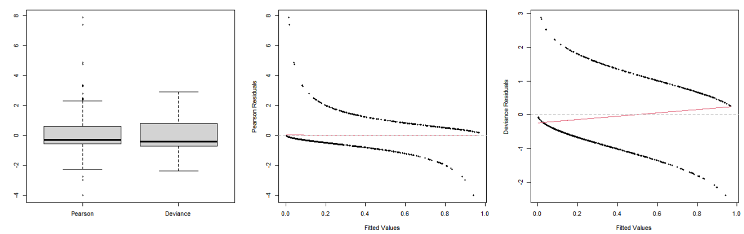
(f)

It is also related with the number of subjects which $Y = 0$ or $Y = 1$.

3. Problem 12

(a)

I used all of predictors without any interaction and non-linear transformations. Build a binomial regression model with link function $\eta = log(\frac{\mu}{1-\mu})$. I used Deviance Residuals and Pearson Residuals to check the goodness of fit.



From those three plots we can know this binomial regression has a good fit with this model.

(b)

The result of GLM (Binomial Model) is showed below:

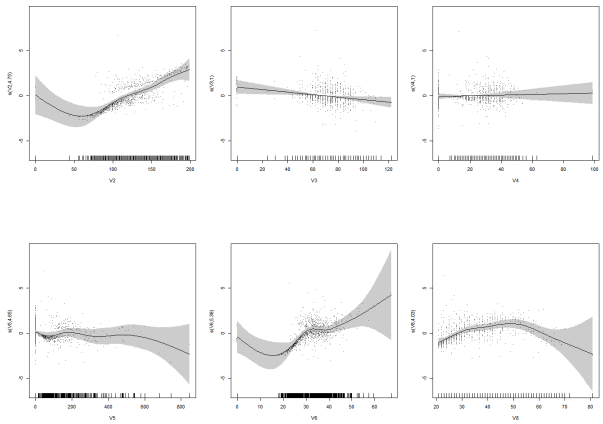
term	estimate	std.error	statistic	p.value
(Intercept)	-8.096098	0.691833	-11.702395	0
V2	0.034609	0.003628	9.540604	0
V3	-0.012522	0.005171	-2.421558	0.015454
V4	0.002723	0.006814	0.399689	0.689386
V5	-0.001121	0.000898	-1.249525	0.211473
V6	0.088973	0.014838	5.996201	0
V8	0.03343	0.008061	4.147112	3.4e-05

From this table we can know the predictors V2, V3, V6 and V8 are all significant under $\alpha = 0.05$. This means those factors are all highly related with response.

The result of GAM is showed below:

term	edf	ref.df	statistic	p.value
s(V2)	4.752839	5.830214	90.068226	0
s(V3)	1.000123	1.000245	6.392987	0.011463
s(V4)	1.000041	1.000081	0.227526	0.633396
s(V5)	4.645374	5.51204	6.248719	0.3872
s(V6)	5.380989	6.51786	37.962591	2e-06
s(V8)	4.034089	5.042671	43.13799	0

From this table we can know the predictors V2, V3, V6 and V8 are all non-linear with response. The fitted additive functions f_j are showed below:



(c)

Yes, I would replace the predictors V4 and V5 by linear function. Because its are all not significant in GAM model. The parametric result table of GPLAM is showed below:

term	df	sumsq	meansq	statistic	p.value
s(V2, 4)	1	99.829822	99.829822	111.82006	0
s(V3, 5)	1	0.150217	0.150217	0.16826	0.681781
V4	1	0.375867	0.375867	0.421011	0.516632
V5	1	2.111236	2.111236	2.36481	0.124522
s(V6, 4)	1	20.18657	20.18657	22.611113	2e-06
s(V8, 5)	1	15.684632	15.684632	17.568462	3.1e-05
Residuals	746.99977	666.900506	0.892772	NA	NA

The non-parametric table are showed below:

term	df	p.value
s(V2, 4)	3	0.000281
s(V3, 5)	4	0.9274613
V4	NA	NA
V5	NA	NA
s(V6, 4)	3	0.0003675
s(V8, 5)	4	7.721e-07

From those result table we can know the factors V2, V6 and V8 are all highly relative with response variable.

4. Problem 13

In this question, we would compare the classification performance of the following four classifiers by 10-fold cross-validation. The predictors which I used are all the factors in PIMA dataset. The estimate of prediction error of each classifier is showed below:

1. (a) Logistic classifier with a linear predictor: 0.1555311
2. (b) Logistic classifier with a quadratic predictor: 0.1575087

3. (c) Fisher's linear discriminant analysis: 0.2264935
4. (d) Quadratic discriminant analysis: 0.2630996

5. Problem 15

(a)

The predictors for all of those four models are

$$\eta_i = \beta_0 + \beta_1 baseline_i + \beta_2 age_i + \beta_3 treatment_{1i}$$

The link functions for Poisson regression model and negative binomial regression model are all $\eta = log(\mu)$.

For the zero-inflated models, if the pmf of the Poisson or Negative Binomial distribution with mean μ is $f(y;\mu)$ with $\mu = g^{-1}(\eta)$, then the zero-inflated model is given by:

$$\begin{aligned}
&P(Y_i = y|X_i) = \alpha_i + (1-\alpha_i)f(y:\mu_i) \quad for \quad y = 0\\
&P(Y_i = y|X_i) = (1-\alpha_i)f(y:\mu_i) \quad for \quad y > 0\\
&logit(\alpha_i) = \tilde{\eta_i} = X_i\gamma
\end{aligned}$$

Here α_i is a probability of a structural zero. Fitting this model requires to jointly fit parameter vector β and γ. The result table of Poisson model is showed below:

term	estimate	std.error	statistic	p.value
(Intercept)	1.942653	0.138286	14.048109	0
baseline	0.022802	0.000831	27.446863	0
age	0.02268	0.004032	5.625471	0
treatment1	-0.147174	0.053523	-2.749753	0.005964

The result table of Negative binomial model is showed below:

term	estimate	std.error	statistic	p.value
(Intercept)	1.869519	0.413262	4.523805	6e-06
baseline	0.030551	0.003505	8.716412	0
age	0.015898	0.012527	1.269113	0.204401
treatment1	-0.172706	0.156138	-1.106112	0.268678

The result table of Zero-Inflate Poisson model is showed below:

term	estimate	std.error	statistic	p.value
(Intercept)	1.8824728	0.1383301	13.609	0
baseline	0.0226070	0.0008316	27.184	0
age	0.0250381	0.0040315	6.211	0.204401
treatment1	-0.1179454	0.0535268	-2.203	0.0276

The result table of Zero-Inflate Negative binomial model is showed below:

term	estimate	std.error	statistic	p.value
(Intercept)	1.78491	0.39640	4.503	0
baseline	0.02979	0.00378	7.881	0
age	0.01962	0.01229	1.597	0.110
treatment1	-0.13618	0.14764	-0.922	0.356

From those four models we can know the predictor baseline is significant in all of four models. The predictor age is only significant in Poisson model. The predictor treatment for positive is significant in Poisson and Zero-inflate Poisson model.

(b)

We can use Vuong non-nested test to test if there exist zero-inflation. The null hypothesis is the two models are indistinguishable. We first use Vuong non-nested test to test if there exist zero-inflation in poisson model. The P-Value for this test is 0.17. It is not significant under significant level of 0.05. So, there does not exist zero-inflation in poisson regression.

Then, We use Vuong non-nested test to test if there exist zero-inflation in negative binomial model. The P-Value for this test is 0.21. It is not significant under significant level of 0.05. So, there does not exist zero-inflation in negative binomial regression.

(c)

Because there is no difference between non-zero-inflate model and zero-inflate model. So, at this moment, we can only consider two non-zero-inflate model, Poisson and Negative Binomial model. However, the overdispersion of Poisson model is 10.35909. It is bigger than 1 with 10 times. The overdispersion of Negative binomial model is 1.15532. It is very similar with 1. So, we should use the negative binomial model as the final model.

(d)

The result table of Negative binomial model is showed below:

term	estimate	std.error	statistic	p.value
(Intercept)	1.869519	0.413262	4.523805	6e-06
baseline	0.030551	0.003505	8.716412	0
age	0.015898	0.012527	1.269113	0.204401
treatment1	-0.172706	0.156138	-1.106112	0.268678

From this table we can know the predictor baseline is significant and the coefficient is positive. This means that under the condition that others factors are the same, the increase of baseline will increase the count of epileptic seizure.

6. Problem 16

I would use Poisson or Negative binomial regression model to predict the count of episodes in the fourth period from the three episode counts of the first three periods.

$$\eta_i = \beta_0 + \beta_1 Count_{i,first} + \beta_2 Count_{i,second} + \beta_3 Count_{i,third}$$
$$Link \quad function : \eta_i = log(\mu_i) = log(E(Y_i|X = x_i))$$

If I want to devise a test to determine whether the influence of all three predictors on the outcome is the same, the null hypothesis and alternative hypothesis are:

- H_0: $\beta_1 = \beta_2 = \beta_3$
- H_1: one of β_j is not equal with each others

At last iteration of IWLS, we can get the estimated $\hat{\beta}$ and weighted matrix $W = W(\hat{\beta})$. By the approximation, we can get

$$\hat{\beta} \sim N(\beta, (X^T W X)^{-1})$$

where the X is the design matrix. By using a linear transformation matrix A we can get

$$A\hat{\beta} \sim N(A\beta, A(X^T W X)^{-1} A^T)$$

where A is a $q \times p$ matrix represents the linear transformation for $\hat{\beta}$, p represents the number of coefficients in the model. In this question, $p = 4$. Under the $H_0 : A\beta = D$, we can have test statistics T:

$$T = (A\hat{\beta} - D)^T [A(X^T W X)^{-1} A^T]^{-1} (A\hat{\beta} - D) \sim \chi^2_q$$

In this question, the null hypothesis is H_0: $\beta_1 = \beta_2 = \beta_3$, for constructing this null hypothesis, we can set

$$A = \begin{bmatrix} 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix}, \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \beta_3 \end{bmatrix}, D = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

This test statistics $T \sim \chi^2_2$.