Generalized Linear Models

Part 1. GLM in General

Linear Model

$$\mu = E(y) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p$$

- The response follows normal distribution
- The relation between the response and the predictors is linear

Generalized Linear Model

$$g(\mu) = \beta_0 + \beta_1 x_1 + ... + \beta_p x_p = x'\beta$$

- The response y follows a distribution in the linear exponential distribution (LED) family
- The relation between the response and the predictors defined by the link function $g(\cdot)$

Linear Exponential Distribution

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

- Examples: Normal, Binomial, Negative Binomial, Poisson, Gamma Distribution, Inverse Gamma.
- The mean and the variance of the distribution are as follows.

$$E(y) = b'(\theta)$$

$$Var(y) = \phi b^{''}(\theta)$$

Prameter Estimation

- Parameters of GLM is $\beta = [\beta_0, \beta_1, ...\beta_p]$
- β can be estimated using the Maximum Likelihood Method

MLE

- Given any data set $(x_1, y_1), (x_2, y_2), ...(x_n, y_n)$ the probability of seeing that data is

$$L(\beta) = f(y_1) \cdot f(y_2) \cdots f(y_n)$$

Case 1: Linear Model

With $E(y) = \mu$

$$g(\mu) = \mu = \beta_0 + \beta_1 x$$

We assume that $Y \sim N(\mu, \sigma^2)$. Let $\epsilon = y - \beta_0 + \beta_1 x$. We have $\epsilon \sim N(0, \sigma^2)$. Therefore

$$y = \beta_0 + \beta_1 x + \epsilon.$$

We see that this is how we define linear model earlier.

Likelihood Function

$$\prod f(y_i) = \prod \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(y_i - \mu_i)^2}{2\sigma^2}}$$
$$= \prod \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}}$$

- The parameters β_0 and β_1 that maximizing the likelihood function are the same as the OLS estimators.

Case 2: Logistic Regression

$$g(\mu) = log\left(\frac{\mu}{1-\mu}\right) = \beta_0 + \beta_1 x, \tag{1}$$

where $Y \sim Bernoulli(p)$, which has the following density distribution

$$f(y) = p^y(1-p)^{1-y}$$

Notice that $\mu = E(y) = p$. Thus, from (1)

$$\mu = \frac{1}{1 - e^{-\beta_0 - \beta_1 x}}$$
$$\implies p = \frac{1}{1 - e^{-\beta_0 - \beta_1 x}}$$

That's why logistic regression is also modeled the probability of $Y = 1$.

Log-Likelihood Function

We have

$$\begin{aligned}
f(y) &= p^y(1-p)^{1-y} \\
\implies \log f(y_i) &= \log p_i^{y_i}(1-p_i)^{1-y_i} \\
&= y_i \log p_i + (1-y_i)\log(1-p_i) \\
&= \log(1-p_i) + y_i \log \frac{p_i}{1-p_i} \\
&= -\log(e^{\beta_0+\beta_1 x}+1) + y_i(\beta_0+\beta_1 x)
\end{aligned}$$

$$\implies \sum \log f(y_i) = \sum -\log(e^{\beta_0+\beta_1 x}+1) + \sum y_i(\beta_0+\beta_1 x)$$

- We differentiate the log likelihood with respect to the parameters and solve for them equal to zero.
- There is no close-form solution. We need to soleve for it numerically.

Example 1

admit	gre	gpa
0	380	3.61
1	660	3.67
1	800	4.00
1	640	3.19
0	520	2.93
1	760	3.00
1	560	2.98
0	400	3.08
1	540	3.39
0	700	3.92
0	800	4.00
0	440	3.22
1	760	4.00
0	700	3.08
1	700	4.00
0	480	3.44
0	780	3.87

	Est.	S.E.	z val.	p
(Intercept)	-4.95	1.08	-4.60	0.00
gre	0.00	0.00	2.54	0.01
gpa	0.75	0.32	2.36	0.02

Standard errors: MLE

$$P(admit = 1) = \frac{1}{1 - e^{-4.949378 + 0.002691 \cdot gre - 0.754687 \cdot gpa}}$$

Coefficients

- The positive coefficient of gpa indicates that increasing gpa will increase the chance of being admitted.
- The same for the gre scores

admit	gre	gpa	predicted_prob
0	380	3.61	0.2310310
1	660	3.67	0.4003934
1	800	4.00	0.5552525
1	640	3.19	0.3057871
0	520	2.93	0.2076762
1	760	3.00	0.3451566
1	560	2.98	0.2326106
0	400	3.08	0.1752786
1	540	3.39	0.2813004
0	700	3.92	0.4731441
0	800	4.00	0.5552525
0	440	3.22	0.2082699
1	760	4.00	0.5285412
0	700	3.08	0.3226841
1	700	4.00	0.4882142
0	480	3.44	0.2569795
0	780	3.87	0.5174912
0	360	2.56	0.1141804

Example 2

income	age	bought_insurance
30	22	0
40	25	0
70	47	1
50	52	0
90	46	1
75	56	1
70	55	0
65	60	1
70	62	1
100	61	1
10	18	0
25	28	0
35	27	0
45	29	0
130	49	1
95	55	1
15	25	1

	Est.	S.E.	z val.	p
(Intercept)	-5.94	2.13	-2.79	0.01
age	0.09	0.05	1.94	0.05
income	0.04	0.03	1.39	0.17

Standard errors: MLE

income	age	bought_insurance	predicted_prob
30	22	0	0.0569696
40	25	0	0.1037916
70	47	1	0.7265557
50	52	0	0.6625367
90	46	1	0.8379541
75	56	1	0.8790796
70	55	0	0.8460245
65	60	1	0.8774682
70	62	1	0.9120940
100	61	1	0.9671907
10	18	0	0.0193308
25	28	0	0.0793765
35	27	0	0.1030927
45	29	0	0.1675142
130	49	1	0.9685957
95	55	1	0.9339869
15	25	1	0.0430395
80	58	1	0.9132951

Case 3. Probit Regression

$Y \sim Bernoulli(p)$

$$
\begin{aligned}
& g(\mu) = \phi^{-1}(\mu) = \beta_0 + \beta_1 x \\
\iff\ & \mu = \phi(\beta_0 + \beta_1 x),
\end{aligned}
$$

where $\phi(z) = \frac{1}{2\pi} \int_{-\infty}^{z} e^{-\frac{u^2}{2}} du$ is the CDF of the standard normal distribution.

income	age	bought_insurance
30	22	0
40	25	0
70	47	1
50	52	0
90	46	1
75	56	1
70	55	0
65	60	1
70	62	1
100	61	1
10	18	0
25	28	0
35	27	0
45	29	0
130	49	1
95	55	1
15	25	1
80	58	1

	Est.	S.E.	z val.	p
(Intercept)	-3.26	1.01	-3.23	0.00
age	0.05	0.03	2.07	0.04
income	0.02	0.01	1.34	0.18

Standard errors: MLE

income	age	bought_insurance	predicted_prob
30	22	0	0.0675621
40	25	0	0.1267136
70	47	1	0.7321767
50	52	0	0.6955872
90	46	1	0.8274191
75	56	1	0.8853669
70	55	0	0.8538991
65	60	1	0.8906097
70	62	1	0.9240656
100	61	1	0.9742320
10	18	0	0.0183095
25	28	0	0.1032072
35	27	0	0.1295963
45	29	0	0.2031544
130	49	1	0.9688680
95	55	1	0.9366018
15	25	1	0.0530884
80	58	1	0.9200541

Case 4. Poisson Regression

$$\log(\mu) = \beta_0 + \beta_1 x$$

where $Y \sim Poisson(\mu)$

We have

$$\begin{aligned} f(y) &= \frac{e^{-\mu}\mu^y}{y!} \\ \implies \log f(y) &= \mu + y \log \mu \log(y!) \\ &= e^{\beta_0+\beta_1 x} + y(\beta_0 + \beta_1 x) - \log(y!) \end{aligned}$$

$$\implies \sum \log f(y_i) = \sum e^{\beta_0+\beta_1 x_i} + y_i(\beta_0 + \beta_1 x_i) - \log(y_i!)$$

- As logistic regression, we need to solve for the derivative equalling zeros using numerical methods.

Example: Predicting Insurance Claims

- ClaimNb Number of claims during the exposure period.
- Exposure The period of exposure for a policy, in years.
- Power The power of the car (ordered categorical).
- CarAge The vehicle age, in years.
- DriverAge The driver age, in years (in France, people can drive a car at 18).
- Brand The car brand divided in the following groups: A- Renaut Nissan and Citroen, B- Volkswagen, Audi, - Skoda and Seat, C- Opel, General Motors and Ford, D- Fiat, E- Mercedes Chrysler and BMW, F- Japanese (except Nissan) and Korean, G- other.
- Gas The car gas, Diesel or regular.
- Region The policy region in France (based on the 1970-2015 classification).
- Density The density of inhabitants (number of inhabitants per km2) in the city the driver of the car lives in.

Model

Prediction

Deviance

- Deviance

$$D = \phi \cdot 2(l_{SAT} - l)$$

- l is the loglikelihood of the data using the model. This is calculated by replacing μ_i by $\hat{\mu_i}$ in the loglikelihood function.
- l_{SAT} is the loglikelihood of the *perfect* model. This is calculated by replacing μ_i by y_i (thus, perfect prediction) in the loglikelihood function.
- ϕ is the scale parameter in the distribution of the response. For logistic regression and Posisson regression $\phi = 1$ and for linear model, $\phi = \sigma^2$

Deviance of Linear Model

- $D = \sum(y_i - \hat{\mu}_i)^2$

Deviance of Logistic Regression

Deviance of Poisson Regression

$$D = 2 \sum \left[y_i \ln \left(\frac{y_i}{\hat{\mu}_i} \right) - (y_i - \hat{\mu}_i) \right]$$

Example

Full Model vs. Reduced Model

- $H_0 : \beta_1 = \beta_2 = \beta_3 = 0$ (Reduced Model)
- $H_\alpha : H_0$ is false. (Full Model)
- $LRT = 2(l_1 - l_0) \sim \chi_{3,\alpha}$ where l_1 and l_0 are the maximum likelihood of the two models.

Example