STAT347: Generalized Linear Models
Lecture 4

Today's topics: Agresi Chapters 4.4, 4.7

- Deviance analysis
- Model diagnosis with residuals
- Example: building a GLM

1 Deviance analysis

Remember that in linear regression, we use R^2, defined as

$$R^2 = 1 - \frac{\sum_i (y_i - \hat{\mu}_i)^2}{\sum_i (y_i - \bar{y})^2} = \frac{\sum_i (\hat{\mu}_i - \bar{y})^2}{\sum_i (y_i - \bar{y})^2}$$

to evaluate how well the model fits the data. We have an analogy in GLM, which is the deviance analysis.

1.1 Definition (more general than the textbook)

Consider density function $f(y;\theta) = e^{\frac{y\theta - b(\theta)}{a(\phi)}} f_0(y;\phi)$ at two values θ_1 and θ_2. Measure the "distance" between two distributions:

$$\begin{aligned} D(\theta_1, \theta_2) = 2\mathbb{E}_{\theta_1}\left\{\log \frac{f(y;\theta_1)}{f(y;\theta_2)}\right\} &= 2\mathbb{E}_{\theta_1}\left\{y(\theta_1 - \theta_2) - b(\theta_1) + b(\theta_2)\right\}/a(\phi) \\ &= 2\left[\mu_1(\theta_1 - \theta_2) - b(\theta_1) + b(\theta_2)\right]/a(\phi) \end{aligned}$$

Remember the 1-to-1 mapping between μ and θ, we also write $D(\mu_1, \mu_2) = D(\theta_{\mu_1}, \theta_{\mu_2})$

- $D(\mu_1, \mu_2) \geq 0$ and the equality holds only when $\mu_1 = \mu_2$
- Generally, $D(\mu_1, \mu_2) \neq D(\mu_2, \mu_1)$
- KL divergence: $D(\mu_1, \mu_2)/2$
- If f is the normal density, then $D(\mu_1, \mu_2) = (\mu_1 - \mu_2)^2/\sigma^2$

Saturated model: imagine the case that we collect an infinite number of covariates, then we can perfectly fit the data and obtain $\hat{\mu}_i = y_i$ for all samples. Then this is called a saturated model.

Deviance between the saturated model (saturated when there is only one observation y): $\hat{\mu} = y$ and another model with μ:

$$\begin{aligned} D(y, \mu) =& 2\left[y(\theta_y - \theta) - b(\theta_y) + b(\theta)\right]/a(\phi) \\ =& -2\log\left[f(y,\theta)/f(y,\theta_y)\right] \end{aligned}$$

With samples $(X_1, y_1), (X_2, y_2), \cdots, (X_n, y_n)$, the total deviance in GLM (the deviance definition in the text book)

$$\begin{aligned} D_+(y, \hat{\mu}) &= \sum_i D(y_i, \hat{\mu}_i) \\ &= -2 \sum_i \log \left[f(y_i, \hat{\theta}_i) / f(y_i, \theta_{y_i}) \right] \end{aligned}$$

This is also called the residual deviance, and compares the estimated GLM model with the saturated model.

Null deviance:

$$\sum_i D(y_i, \bar{y})$$

where $\bar{y} = \sum_i y_i / n$. The null deviance compares the null model ($\mu_i \equiv \mu$) with the saturated model.

1.2 Deviance analysis for nested models

Let $\beta = \begin{pmatrix} \beta^{(1)} \\ \beta^{(2)} \end{pmatrix}$ where $\beta^{(1)} \in \mathbb{R}^{p_1}$ and $X = \begin{pmatrix} X^{(1)} & X^{(2)} \end{pmatrix}$.

We call $\mathcal{M}^{(1)}$ with

$$g(\mu_i) = X^{(1)} \beta^{(1)}$$

a nested model of the full model $\mathcal{M}$ where

$$g(\mu_i) = X\beta.$$

Let $\hat{\beta}^{(1)}$ be the MLE solution of the model $\mathcal{M}^{(1)}$ and $\hat{\mu}^{(1)}$ be the corresponding estimated expectations of y in the fitted model.

Then,

$$D_+(\hat{\mu}, \hat{\mu}^{(1)}) = D_+(y, \hat{\mu}^{(1)}) - D_+(y, \hat{\mu}) = -2 \left[L(\hat{\beta}^{(1)}) - L(\hat{\beta}) \right]$$

is the likelihood ratio between two models.

- Test for $H_0 : \beta^{(2)} = 0$. Under H_0,

$$D_+(y, \hat{\mu}^{(1)}) - D_+(y, \hat{\mu}) \to \mathcal{X}^2_{p-p_1}$$

- Compare with the null model, we can also define "R^2" in GLM:

$$1 - \frac{D_+(y, \hat{\mu})}{\sum_i D(y_i, \bar{y})}$$

1.3 Model comparison with deviance analysis table

Say we partition our covariates as

$$X = (1, X_{(1)}, X_{(2)}, \cdots, X_{(J)})$$

and $X_{(j)} \in \mathbb{R}^{d_j}$. We can sequentially add each partition of covariates into the model (in some pre-determined order) and understand each partition's "relative contribution" with a deviance analysis table.

Define the following quantities:

- $\hat{\beta}^{(j)}$ is the MLE solution of the GLM model with covariates $X^{(j)} = (1, X_{(1)}, X_{(2)}, \cdots, X_{(j)})$
- $\hat{\mu}^{(j)}$ is the corresponding vector of expectations of $y = (y_1, \cdots, y_n)$ in the fitted model.

Then the deviance analysis table is shown in Table 1.
The difference of two residual deviances

$$D_+(y, \hat{\mu}^{(j-1)}) - D_+(y, \hat{\mu}^{(j)}) = 2L(\hat{\beta}^{(j)}) - 2L(\hat{\beta}^{(j-1)})$$

so that we can use the likelihood ratio test.

Model	twice log-likelihood	residual deviance	difference	df	Compare with
$\hat{\beta}^{(0)}$ (null)	$2L(\hat{\beta}^{(0)})$	$D_+(y, \hat{\mu}^{(0)}) = \sum_i D(y_i, \bar{y})$			
$\hat{\beta}^{(1)}$	$2L(\hat{\beta}^{(1)})$	$D_+(y, \hat{\mu}^{(1)})$	$D_+(y, \hat{\mu}^{(0)}) - D_+(y, \hat{\mu}^{(1)})$	d_1	$\chi^2_{d_1}$
$\hat{\beta}^{(2)}$	$2L(\hat{\beta}^{(2)})$	$D_+(y, \hat{\mu}^{(2)})$	$D_+(y, \hat{\mu}^{(1)}) - D_+(y, \hat{\mu}^{(2)})$	d_2	$\chi^2_{d_2}$
$\vdots$					
$\hat{\beta}^{(J)}$	$2L(\hat{\beta}^{(J)})$	$D_+(y, \hat{\mu}^{(J)})$	$D_+(y, \hat{\mu}^{(J-1)}) - D_+(y, \hat{\mu}^{(J)})$	d_J	$\chi^2_{d_J}$

Table 1: Deviance analysis table.

2 Model checking with the residuals

As in the linear models, we can examine the residuals to help us check whether a model fits poor or not, and whether there are any outliers in the observations.
Three types of residuals:

- Pearson residual:
$$e_i = \frac{y_i - \hat{\mu}_i}{\sqrt{v(\hat{\mu}_i)}}$$
where $v(\hat{\mu}_i) = \widehat{\text{Var}}(y_i)$. For instance, if $y_i \sim \text{Poisson}(\mu_i)$ then $v(\hat{\mu}_i) = \hat{\mu}_i$. As we have shown in Lecture 2, in general $v(\hat{\mu}_i) = b''(\hat{\theta}_i)a(\hat{\phi})$.
- Deviance residual:
$$d_i = \sqrt{D(y_i, \hat{\mu}_i)} \times \text{sign}(y_i - \hat{\mu}_i)$$
For instance, for the Gaussian linear model, $D(y_i, \hat{\mu}_i) = (y_i - \hat{\mu}_i)^2/\sigma^2$, and the deviance residual is the same as the Pearson residual. As a rule of thumb, an observation is fitted poorly by the GLM model if $|d_i| > 2$.
- As in the linear models, the mean of e_i is typically smaller than 1 as $\hat{\mu}_i$ is estimated. After some calculations (see Chapter 4.4.5), one can compute a more accurate variance of $y_i - \hat{\mu}_i$.

  Standardized residual:
$$r_i = \frac{e_i}{\sqrt{1 - \hat{h}_{ii}}}$$
where h_{ii} is the ith diagonal element of the H_W defined equation (4.19) of the Agresti chapter 4.4.5.

3 Data examples

Please check the R notebook 2.

Next time: Chapter 4.5, Chapter 5.1 - 5.2, binary data model, application scenarios