Signal-to-noise Ratio in Linear Regression and Generalized Linear Models

Zihan Zhu

July 2023

1 Introduction

The signal-to-noise ratio (SNR) is a measure used in various fields. It compares the strength of a desired signal to the level of background noise present in a system. In simple terms, the signal-to-noise ratio quantifies the quality of a signal by determining how much of the received or measured data is the actual useful signal and how much is unwanted noise. A high SNR indicates that the signal is strong and distinguishable from the noise, while a low SNR suggests that the signal is weak and may be difficult to extract from the noise. Let S be the random variable of signal and N be the random variable of noise. A simple formula of SNR is

$$\text{SNR} = \frac{E[S^2]}{E[N^2]} \tag{1}$$

If the signal is a quantity without randomness, SNR is

$$\text{SNR} = \frac{S^2}{E[N^2]} \tag{2}$$

2 SNR in Linear Regression

Consider multiple linear regression model with p predictors

$$Y = X\beta + \epsilon, \epsilon \sim N(0, \sigma^2 I_n) \tag{3}$$

Let $J_{n,m}$ be $n \times m$ matrix with all elements being 1. The SNR is

$$\text{SNR} = \frac{(\bar{Y} J_{n,1} - X\beta)'(\bar{Y} J_{n,1} - X\beta)}{(Y - X\beta)'(Y - X\beta)} \tag{4}$$

Since β is unknown, we can use OLS of β instead to get an estimator of SNR:

$$\hat{\text{SNR}} = \frac{(\bar{Y} J_{n,1} - X\hat{\beta})'(\bar{Y} J_{n,1} - X\hat{\beta})}{(Y - X\hat{\beta})'(Y - X\hat{\beta})} \tag{5}$$

Note that the denominator is sum of squared error, while the numerator the the regression sum of squares

$$\text{SSE} = (Y - X\hat{\beta})'(Y - X\hat{\beta}), \text{SSR} = (\bar{Y} J_{n,1} - X\hat{\beta})'(\bar{Y} J_{n,1} - X\hat{\beta}) \tag{6}$$

Hence, the estimator of SNR can be interpreted as the regression mean squares,

$$\text{MSR} = \frac{\text{SSR}}{\text{SSE}} = \frac{(\bar{Y} J_{n,1} - X\hat{\beta})'(\bar{Y} J_{n,1} - X\hat{\beta})}{(Y - X\hat{\beta})'(Y - X\hat{\beta})} \tag{7}$$

which means we can estimate the SNR of linear regression via ANOVA.

3 ANOVA for GLM

Consider GLM defined by exponential dispersion family and link function g,

$$
\begin{aligned}
&Y_i \sim \mathrm{EDM}(\mu_i, \phi/w_i) \\
&\eta_i = g(\mu_i) = \beta_1 x_{1,i} + \cdots + \beta_p x_{p,i}
\end{aligned}
\tag{8}
$$

assuming $x_{1,i} = 1$. The density of Y_i is

$$
f(y_i|\mu_i, \phi) = b(y_i, \phi/w_i) \exp\left(-\frac{d(y_i, \mu_i)}{2\phi/w_i}\right) \tag{9}
$$

where $d(y_i, \mu_i)$ is the unit deviance and w_i is the prior weight. Total deviance of the model is

$$
D(y, \hat{\mu}) = \sum_{i=1}^{n} w_i d(y_i, \hat{\mu}_i) \tag{10}
$$

If we regard the unit deviance as the generalization of $(y_i - x_i'\beta)^2$, then the total deviance can be regarded as the generalization of sum of squared residuals (SSE). However, the the contribution of explanatory variables can't be calculated directly from the full model. Hence, we consider a reduced model

$$
\begin{aligned}
&Y_i \sim \mathrm{EDM}(\mu_i, \phi/w_i) \\
&\eta_i = g(\mu_i) = 0
\end{aligned}
\tag{11}
$$

Let $\hat{\mu}_0$ be the fitted value of the reduced model. And the deviance reduced by regression is

$$
D(y, \hat{\mu}_0) - D(y, \hat{\mu})
$$

Above, the contribution of regression is

$$
\frac{D(y, \hat{\mu}_0) - D(y, \hat{\mu})}{D(y, \hat{\mu})} \tag{12}
$$

which can be used as the estimator of SNR of GLM.

4 Example: The SNR of Poisson Regression

For Poisson regression,

$$\begin{aligned} f(y_i|\mu_i) &= \frac{\mu_i^{y_i} e^{-\mu_i}}{y_i!} \\ &= \frac{1}{y_i!} \exp\left(y_i \log \mu_i - \mu_i\right) \end{aligned} \tag{13}$$

the canonical parameter is $\theta_i = \log \mu_i$ and cumulant function is $\kappa(\theta_i) = \mu_i = e^{\theta_i}$, Let

$$t(y_i, \mu_i) = y_i \log \mu_i - \mu_i$$

The unit deviance is

$$d(y_i, \mu_i) = 2[t(y_i, y_i) - t(y_i, \mu_i)] = 2\left(y_i \log \frac{y_i}{\mu_i} - (y_i - \mu_i)\right) \tag{14}$$

Consider canonical link, $\eta_i = g(\mu_i) = \log \mu_i = 0$, then $\mu_i = 1$. Hence, in the reduced model,

$$d(y_i, \mu_0) = 2\left(y_i \log y_i - (y_i - 1)\right) \tag{15}$$

Above,

$$\hat{\text{SNR}} = \frac{\sum_{i=1}^n \left(y_i \log \frac{y_i}{\hat{\mu}_i} - (y_i - \hat{\mu}_i)\right)}{\sum_{i=1}^n \left(y_i \log y_i - (y_i - 1)\right)} \tag{16}$$