General Linear Model

$$Y \sim N(\underbrace{\beta X}_{E(Y)}, \sigma^2)$$

$$\hat{Y} = \hat{\beta} X \rightarrow$$ GLM

MLE

Select $\hat{\beta}$ to minimize L_2

$$L_2(\hat{\beta}) = \| Y - \hat{Y} \|^2$$

$$Y = \hat{Y}_1 + \varepsilon_1$$

$$Y = \hat{Y}_2 + \varepsilon_2$$ — ova

compare models

Maximum Likelihood Estimator

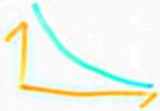
Select model parameters $\hat{\beta}$ that maximize the Likelihood of the model.

$$\mathcal{L}(\hat{\beta}) = P(\text{data} \mid \beta = \hat{\beta})$$

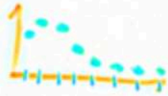
The likelihood of a model is the probability of observing the data if the model is correct.

General$_{ized}$ Linear Models

$$Y \sim N(\pi, \sigma^2) \longrightarrow y \in \mathbb{R} \rightarrow g(\pi) = \pi$$



$$Y \sim exp(\pi) \longrightarrow y \in \mathbb{R}^+ \rightarrow g(\pi) = \pi^{-1}$$



$$Y \sim poisson(\pi) \longrightarrow y \in \mathbb{Z}^+ \rightarrow g(\pi) = \ln(\pi)$$

$$Y \sim bernoulli(\pi) \longrightarrow y \in [0,1] \rightarrow g(\pi) = \ln\left(\frac{\pi}{1-\pi}\right)$$

$$g(\pi) = \beta X \qquad \hat{Y} = g^{-1}(\hat{\beta} X)$$

Logistic Regression

$$Y \sim \text{bernoulli}(\underbrace{\pi}_{E(Y)=P(Y=1)}) \quad Y \in [0,1]$$

$$\hat{Y} = \hat{\pi} = g^{-1}(\hat{\beta}X) \longrightarrow \text{Logistic MLE}$$

Select $\hat{\beta}$ to minimize the Cross Entropy of Y and $\hat{Y}$

$$H(Y,\hat{Y}) = -\sum_{i}^{N} P(Y_i) \cdot \log_2(P(\hat{Y}_i))$$

$$g(\pi) = \underbrace{\ln\left(\frac{\pi}{\pi - 1}\right)}_{\text{log odds or logit}}$$

Cross Entropy measures the information lost by using $\hat{Y}$ in place of Y.

A logistic MLE minimizes $H(Y,\hat{Y})$

Cross Entropy measures the negative log likelihood of a logistic regression model.

$$Y = 1 \quad \hat{Y} = 0.8$$

$$H(Y,\hat{Y}) = -\left[1 \cdot \log_2(0.8) + 0 \cdot \log_2(0.2)\right] = 0.32$$

$$Y = 0 \quad \hat{Y} = 0.8$$

$$H(Y,\hat{Y}) = -\left[0 \cdot \log_2(0.8) + 1 \cdot \log_2(0.2)\right] = 2.32$$

Drop in Deviance Testing

$$\text{deviance} = C - 2\cdot\ln(\mathcal{L}_{MAX})$$

$$\Delta\,\text{deviance} = 2\cdot\ln\left(\frac{\mathcal{L}_{MAX_F}}{\mathcal{L}_{MAX_R}}\right)$$

Δ deviance measures the <u>drop in deviance</u> from adding predictors to a reduced model (R) to obtain a full model (F)

$$p\text{-value} = P(\chi^2_{df} \geq \Delta\,\text{deviance})$$

df = # of predictors added to the reduced model.

Δ deviance tests are also called Likelihood Ratio Tests (LRT)

Deviance is to sum of squared residuals as Maximum Likelihood Estimation is to ordinary least squares estimation.