linear model w/ Gaussian error terms

$$y_i \mid x_i \overset{ind}{\sim} N(x_i^T \beta_*, \sigma^2)$$

(binary) logistic regression

$$y_i \mid x_i \overset{ind}{\sim} Ber\left(\frac{e^{x_i^T \beta_*}}{1+e^{x_i^T \beta_*}}\right)$$

generic generalized linear model (GLM)

$$y_i \mid x_i \overset{ind}{\sim} \underbrace{\boxed{F}}_{\text{some "nice" distribution}}\left(\underbrace{\boxed{\gamma}}_{\text{response function}}(x_i^T \beta_*)\right)$$

Poisson regression:

motivation: regression / supervised learning problems in which y_i is a count

i.e. $y_i \in \{0, 1, 2, \dots\}$.

$$y_i \mid x_i \sim Poi(e^{x_i^T \beta_*})$$

Recall $Y \sim Poi(\lambda)$, then pmf of Y is

$$P(Y=c) = \frac{\lambda^c e^{-\lambda}}{c!}$$

General design pattern:

F is an exponential family

γ is the inverse of the gradient of log partition function.

Multi-class logistic regression

$$y_i \mid x_i \overset{ind}{\sim} Cat\left(\begin{bmatrix} \frac{e^{x_i^T \beta_1^*}}{e^{x_i^T \beta_1^*}+\dots+e^{x_i^T \beta_K^*}} \\ \vdots \\ \frac{e^{x_i^T \beta_K^*}}{e^{x_i^T \beta_1^*}+\dots+e^{x_i^T \beta_K^*}} \end{bmatrix}\right)$$

Aside: $Y \sim Cat(\pi)$, then

$Y \in \{y_1, y_2, \dots, y_K\}$

$$P(Y \cdot) = \begin{cases} \pi_1 & \text{if } Y = y_1 \\ \vdots & \\ \pi_K & \text{if } Y = y_K \end{cases}, \quad \pi = \begin{bmatrix} \pi_1 \\ \vdots \\ \pi_K \end{bmatrix}$$

$$= \pi_1^{\mathbb{1}\{Y=y_1\}} \pi_2^{\mathbb{1}\{Y=y_2\}} \cdots \pi_K^{\mathbb{1}\{Y=y_K\}}$$

Ex: (MLE for multiclass logistic)

$$L(\beta_1, \dots, \beta_K) = \prod_{i=1}^n p(y_i \mid x_i; \beta_1, \dots, \beta_K)$$

$$= \prod_{i=1}^n \left(\frac{e^{x_i^T \beta_1}}{\dots}\right)^{\mathbb{1}\{y_i = y_1\}} \left(\frac{e^{x_i^T \beta_2}}{\dots}\right)^{\mathbb{1}\{y_i = y_2\}}$$

$$= \prod_{i=1}^{n} \left(\frac{\cdots}{\sum_{l=1}^{K} e^{x_i^T \beta_l}} \right) \cdots \left(\frac{\cdots}{\sum_{l=1}^{K} e^{x_i^T \beta_l}} \right) \cdots$$

$$\cdots \left(\frac{e^{x_i^T \beta_K}}{\sum_{l=1}^{K} e^{x_i^T \beta_l}} \right)^{\mathbb{1}\{y_i = y_K\}}.$$

$$= \prod_{i=1}^{n} \left(\prod_{m=1}^{K} \left(\frac{e^{x_i^T \beta_m}}{\sum_{l=1}^{K} e^{x_i^T \beta_l}} \right)^{\mathbb{1}\{y_i = y_m\}} \right)$$