How to estimate params for logistic regression?

Recall $$p(x) = \frac{e^{\beta_0+\beta_1 x}}{1+e^{\beta_0+\beta_1 x}}, \text{ or}$$

$$g(p(x)) = \beta_0 + \beta_1 x \quad , \text{ where } g(u) = \log\left(\frac{u}{1-u}\right)$$

Params to be estimated: β_0, β_1

$p(x) = \mathbb{P}[Y=1 \mid X=x]$ is NOT pdf

Data to be used for MLE purpose?

Labels $y_1, \dots, y_n$

$$\{(\underbrace{x_i}_{\text{predictor}}, \underbrace{y_i}_{\text{Label}})\}_{i=1}^{n}$$

$$L(\theta) = \prod_{i=1}^{n} \mathbb{P}[Y=y_i \mid X=x_i]$$

Want to maximize the likelihood of labels y_i given the predictors x_i.

$$L(\theta) = \prod_{i=1}^{n} p(x_i)^{y_i}[1-p(x_i)]^{1-y_i} \quad , \quad \underset{\sim}{\theta} \overset{\text{def}}{=} (\beta_0, \beta_1)$$

$$\ln L(\theta) = \sum_{i=1}^{n} y_i \ln p(x_i) + (1-y_i)\ln(1-p(x_i))$$

$$= \sum \ln(1-p(x_i)) + \sum y_i \underbrace{\ln \frac{p(x_i)}{1-p(x_i)}}_{\beta_0+\beta_1 x_i}$$

$$= \sum -\ln(1+e^{\beta_0+\beta_1 x_i}) + \sum y_i(\beta_0+\beta_1 x_i)$$

$$\frac{\partial \ln L(\theta)}{\partial \beta_1} = \sum_{i=1}^{n} \left[y_i - \underbrace{p(x_i; \beta_0, \beta_1)} \right] x_i \overset{\text{set}}{=} 0$$

nonlinear, needs to be solved using Iterative Reweighted LS.

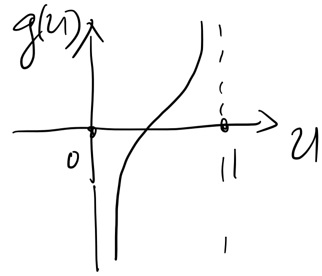
For Linear Regression:

$$\min_{\beta_1} \sum_{i=1}^{n} (y_i - \beta_1 x_i)^2$$

$$\frac{\partial}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_1 x_i) x_i \overset{\text{set}}{=} 0$$

$$\Rightarrow \hat{\beta}_1 = \frac{\sum x_i y_i}{\sum x_i^2}$$

Why logit? $g(u)=\log\left(\frac{u}{1-u}\right)$



Historically, "probit" was used

$g = \Phi^{-1}$ inverse of CDF for Gaussian

For Linear regression, $g(\mu_i)=\beta_0+\beta_1 x_i$, $g(\cdot)$ is identity

$\mu_i \overset{def}{=} \mathbb{E}[Y_i | X=x_i]$

For logistic reg, $g(\mu_i)=\beta_0+\beta_1 x_i$, $g(u)=\log\frac{u}{1-u}$

For Poisson dist, $g(\mu_i)=\beta_0+\beta_1 x_i$, $g(u)=\ln(u)$

$Y_i \in \{0,1,\dots\}$

link func.

Exponential Family Dist:

$$f(y;\underset{\text{natural param}}{\theta}) = \exp\left[\frac{y\theta-b(\theta)}{\underbrace{a(\phi)}_{\text{scale param}}}+c(y,\phi)\right],$$ where ϕ is fixed

Bernoulli: $\mathbb{P}[Y=y]=p^y(1-p)^{1-y}=\exp\left[\ln(p^y)+\ln(1-p)^{1-y}\right]$

$= \exp\left[y\cdot\underset{\theta}{\boxed{\ln\left(\frac{p}{1-p}\right)}}+\ln(1-p)\right]$

Poisson: $\mathbb{P}[Y=y]=\frac{\mu^y e^{-\mu}}{y!}=\exp\left[y\cdot\underset{\theta}{\boxed{\ln\mu}}-\mu-\ln(y!)\right]$

Gaussian: $f_Y(y) = \frac{1}{\sqrt{2\pi}\sigma} e^{-(y-\mu)^2/2\sigma^2}$ (σ^2 fixed, known)

$$= \exp\left[-\ln(\sqrt{2\pi}\sigma) - \frac{(y-\mu)^2}{2\sigma^2}\right]$$

$$= \exp\left[\frac{y \cdot \cancel{2}\overset{\theta}{\boxed{\mu}} - \cdots \cancel{2}}{\cancel{2}\underbrace{\sigma^2}_{\leftarrow \text{ scale param}}} + c(y, \phi)\right]$$

$g(\cdot)$ such selected are called "canonical link func"

Exp family dist can use the same method, i.e., MLE to find the best params:

$$\log l(\underset{\sim}{\beta}) = \log\left(\prod_{i=1}^{n} f(y_i; \beta)\right)$$

$$= \sum_{i=1}^{n}\left[\frac{y_i\theta_i - b(\theta_i)}{a(\phi)} + c(y_i, \phi)\right]$$

$\frac{\partial}{\partial \beta_i} \log(\underset{\sim}{\beta}) = 0$ for all i.

Need IRWLS. Details in McCulloh 5.4.

[Generalized Linear Model (GLM)]