Logit Model Estimation.

(Maximum Likelihood Method).

$Pr(Y/X)$ ← probability → Yes / NO. } Binary → 1 / 0

Let's pick one of the classes and call it 1 and other is 0. In this case Y becomes an indicator variable and you can say that

$$\boxed{\Sigma, \Pi}$$

$$Pr(Y=1) = E(Y)$$

in terms of conditional probability, $Pr(Y=1 / X=x) = E[Y / X=x]$

[Assume that $Pr(Y=1/X=x) = p(x;\theta)$

then the conditional likelihood function is,

$$\prod_{i=1}^{n} Pr(Y=y_i \mid X=x_i) = \prod_{i=1}^{n} p(x_i;\theta)^{y_i}(1-p(x_i;\theta))^{1-y_i}$$

Recall that in a sequence of Bernoulli trials $y_1, y_2, \dots, y_n$ where there is a constant probability of success p, the likelihood is

$$\prod_{i=1}^{n} p^{y_i}(1-p)^{1-y_i}$$

LOGISTIC REGRESSION:

LOGISTIC REGRESSION.

So we have a binary output variable Y, and we want to model the conditional probability $Pr(Y=1|X=x)$ as a function of x; any unknown parameters in the function are to be estimated by maximum likelihood.

How can we use linear regression then?

(1) The most obvious idea is to let $p(x)$ be a linear function of $\underline{x}$. Every increment of component of x would add or subtract to the probability. Here p must be 0 and 1.

(2) The next idea is to take $\underline{\log p(x)}$ be a linear function of x, so that changing an input variable multiplies the probability by a fixed amount.

(3) The easiest modication of $\log p$ is logistic (or logit) transformation, $\log\left(\frac{p}{1-p}\right)$

Formally the model logistic regression model is that,

$$\log \frac{p(x)}{1-p(x)} = \beta_0 + x \cdot \beta$$

Solving for p, This gives

$\beta_0 + x \cdot \beta$

solving for p, this gives

$$p(x;b,w) = \boxed{\frac{e^{\beta_0 + x\cdot\beta}}{1+e^{\beta_0+x\beta}}} = \frac{1}{1+e^{-(\beta_0+x\beta)}}$$

we should predict $Y=1$ when $p \geq 0.5$

& $Y=0$ when $p<0.5$

Likelihood function for Logistic Regression:

Because Logistic regression predicts probabilities, rather than just classes, we can fit it using likelihood. For each data point, we have a vector of features, x_i and an observed class, y_i. The probability of that class was either

p, if $y_i=1$ or $1-p$ if $y_i=0$

The Likelihood is then,

$$L(\beta_0,\beta) = \prod_{i=1}^{n} p(x_i)^{y_i}(1-p(x_i))^{1-y_i} = p(x_i)^{\sum_{i=1}^{n} y_i} \cdot (1-p(x_i))^{\Sigma(1-y_i)}$$

Let us take log on both sides, then the log-likelihood turns as follows.

$$\log L(\beta_0,\beta) = \sum_{i=1}^{n} y_i \log p(x_i) + \Sigma(1-y_i)\log(1-p(x_i))$$

$$= \sum^{n} \ldots \log p(x_i) + \sum^{n} \log (1-p(x_i)$$

$$= \underbrace{\sum_{i=1}^{n} y_i}\cdot \log p(x_i) + \sum_{i=1}^{n} \log (1-p(x_i)) - \underbrace{\sum_{i=1}^{n}} y_i \log (1-p(x_i))$$

$$= \sum \log (1-p(x_i)) + \sum y_i \left[\log p(x_i) - \log(1-p(x_i))\right]$$

$$= \sum \log (1-p(x_i)) + \sum y_i \left(\log \frac{p(x_i)}{1-p(x_i)}\right)$$

$$= \sum \log (1-p(x_i)) + \sum y_i \, (\beta_0 + x_i \cdot \beta)$$

$$\text{Log } L(\beta_0, \beta) = \sum_{i=1}^{m} -\log\left(1 + e^{\beta_0 + x_i \beta}\right) + \sum_{i=1}^{m} y_i (\beta_0 + x_i \beta)$$

$\hookrightarrow$ Log-Likelihood function (LLF).

To find the maximum Likelihood estimates we'd differentiate the log likelihood w.r.t parameters,

$$\frac{\partial \text{ Log } L(\beta_0, \beta)}{\partial \beta_j} = -\sum \frac{1}{\left(1+e^{\beta_0 + x_i \beta}\right)} \left(e^{\beta_0 + x_i \beta} x_{ij}\right) + \sum y_i x_{ij}$$

$$= \sum \left(y_i - p(x_i; \beta_0, \beta)\right) x_{ij}$$