Question 1

Monday, 19 February 2024 10:21 PM

Logistic Regression

Given N pairs of data points : $(x_i, c_i), i = 1, \dots N$

$x_i \rightarrow$ feature vector

$c_i \rightarrow$ binary class label (0 or 1)

$$y_i = \beta_1^T x_i + \beta_0 \quad \text{--- (1)}$$

also $y_i = \log\left(\frac{p}{1-p}\right)$ where p is probability ($c_i = 1$)

$\Rightarrow$ $1-p$ is probability for $c_i = 0$

Likelihood function $= \prod_{i=1 \dots N} p(x_i)^{c_i} (1 - p(x_i))^{(1-c_i)}$

where $p(x_i)$ is probability of x_i is from class 1.

We need to find the Log-likelihood

We know from above

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

$$\Rightarrow e^{y_i} = \frac{p(x_i)}{1-p(x_i)} \quad \text{(Taking exponent both sides)}$$

$$\Rightarrow (1-p(x_i))e^{y_i} = p(x_i)$$

$$\Rightarrow e^{y_i} = e^{y_i} p(x_i) + p(x_i)$$

$$\Rightarrow p(x_i) = \frac{e^{y_i}}{1+e^{y_i}} \quad \text{--- (2)}$$

$$\text{also, } 1 - p(x_i) = \frac{1}{1+e^{y_i}} \quad \text{--- (3)}$$

Log-likelihood = log (likelihood func^n)

$$= \log\left(\prod_{1,\dots,N} (p(x_i))^{c_i} \times (1 - p(x_i))^{(1-c_i)}\right)$$

$$= \sum_{i=1}^{N} \left[c_i \log(p(x_i)) + (1-c_i)\log(1-p(x_i))\right]$$

Substituting values from 2 & 3

$$\text{log-Likelihood} = \sum_{i=1}^{N}\left[C_i \log\left(\frac{e^{y_i}}{1+e^{y_i}}\right) + (1-C_i)\log\left(\frac{1}{1+e^{y_i}}\right)\right]$$

$$= \sum_{i=1}^{N}\Big[C_i \log(e^{y_i}) - C_i \log(1+e^{y_i}) + (1-C_i)\log 1$$

$$- (1-C_i)\log(1+e^{y_i})\Big]$$

$$= \sum_{i=1}^{N}\Big[C_i y_i - \cancel{C_i \log(1+e^{y_i})} + (1-C_i)\times 0 - \log(1+e^{y_i})$$

$$+ \cancel{C_i \log(1+e^{y_i})}\Big]$$

$$= \sum_{i=1}^{N}\left(C_i y_i - \log(1+e^{y_i})\right)$$

Since $y_i = \beta_1^T x_i + \beta_0$ (from 1

$$\Rightarrow \underset{(\mathcal{L}(\beta))}{\text{log-likelihood}} = \sum_{i=1}^{N}\left[C_i(\beta_1^T x_i + \beta_0) - \log\left(1 + \exp(\beta_1^T x_i + \beta_0)\right)\right]$$

Hence Proved

Gradient Computation

For the log-likelihood function

$$\mathcal{L}(\beta) = \sum_{i=1}^{N}\left[C_i(\beta_1^T x_i + \beta_0) - \log\left(1 + \exp(\beta_1^T x_i + \beta_0)\right)\right]$$

we will compute the partial derivatives with respect to both β_1 & β_0

where $\beta_1 \rightarrow$ Intercept & $\beta_0 \rightarrow$ coefficient of slope

& $\nabla_\beta L(\beta) = \left[\frac{\partial L}{\partial \beta_0}, \frac{\partial L}{\partial \beta_1}\right]^T$

Partial derivative w.r.t. β_0

$$\frac{\partial L}{\partial \beta_0} = \sum_{i=1}^{N} \left\{ C_i \frac{\partial}{\partial \beta_0} (\beta_1^T x_i + \beta_0) - \frac{\partial}{\beta_0} \left[\log \left(1 + \exp(\beta_1^T x_i + \beta_0) \right) \right] \right\}$$

$$= \sum_{i=1}^{N} \left\{ C_i - \frac{1}{1 + \exp(\beta_1^T x_i + \beta_0)} \times \frac{\partial}{\partial \beta_0} \left[1 + \exp(\beta_1^T x_i + \beta_0) \right] \right\}$$

$$= \sum_{i=1}^{N} \left[C_i - \frac{\exp(\beta_1^T x_i + \beta_0)}{1 + \exp(\beta_1^T x_i + \beta_0)} \times \frac{\partial}{\partial \beta_0} (\beta_1^T x_i + \beta_0) \right]$$

$$= \sum_{i=1}^{N} \left[C_i - \frac{\exp(\beta_1^T x_i + \beta_0)}{1 + \exp(\beta_1^T x_i + \beta_0)} \right] \quad \text{— } 4$$

Partial derivative w.r.t. β_1

$$\frac{\partial L}{\partial \beta_1} = \sum_{i=1}^{N} \left[C_i \frac{\partial}{\partial \beta_1} (\beta_1^T x_i + \beta_0) - \frac{\partial}{\partial \beta_1} \log \left(1 + \exp(\beta_1^T x_i + \beta_0) \right) \right]$$

$$= \sum_{i=1}^{N} \left[C_i x_i - \frac{\exp(\beta_1^T x_i + \beta_0)}{1 + \exp(\beta_1^T x_i + \beta_0)} \times \frac{\partial}{\partial \beta_1} (\beta_1^T x_i + \beta_0) \right]$$

(Skipped one step, which is similar to above)

$$= \sum_{i=1}^{N} \left[C_i x_i - \frac{x_i \times \exp(\beta_1^T x_i + \beta_0)}{1 + \exp(\beta_1^T x_i + \beta_0)} \right] \quad \text{— } 5$$

$\Rightarrow$ We can say that

$$\nabla_\beta L(\beta) = \nabla_\beta (\text{log-likelihood})$$

$$= \left[\frac{\partial L}{\partial \beta_0}, \frac{\partial L}{\partial \beta_1} \right]^T$$

And we can use the equations from 4 & 5

Ans