classmate
Date
Page

ODSM Assignment - II

Part B : Mathematical formulation

B.1 Derivation for Negative Log-Likelihood

$$P(y_i = 1|x_i) = \sigma(\beta_0 + \beta^T x_i)$$

→ For binary classification

$$P(y_i = 0|x_i) = 1 - P(y_i = 1|x_i)$$

$$\Rightarrow \quad P(y_i|x_i) = P(y_i = 1|x_i)^{y_i} \cdot P(y_i = 0|x_i)^{(1-y_i)}$$

⇒ ~~to~~ Likelihood for given logistic regression with N datapoints is

$$\mathcal{L}(\beta_0, \beta) = \prod_{i=1}^{N} P(y_i|x_i) = \prod_{i=1}^{N} \left[\sigma(\beta_0 + \beta^T x_i)^{y_i} \cdot \{1 - \sigma(\beta_0 + \beta^T x_i)\}^{(1-y_i)} \right]$$

→ Taking log on both sides

$$\Rightarrow \quad \log(\mathcal{L}(\beta_0, \beta)) = \sum_{i=1}^{N} \left[y_i \log(\sigma(\beta_0 + \beta^T x_i)) + (1 - y_i) \cdot \log(1 - \sigma(\beta_0 + \beta^T x_i)) \right]$$

$$= \sum_{i=1}^{N} \left[y_i \log(p_i) + (1-y_i) \log(1-p_i) \right]$$

$$\Rightarrow \quad NLL = -\log(\mathcal{L}(\beta_0, \beta))$$

$$= -\sum_{i=1}^{N} \left[y_i \log(p_i) + (1-y_i) \log(1-p_i) \right]$$

⇒ Hence derived!

B. 2 Derivation for gradient

(i) Gradient w.r.t. β_0 :

→ let $z_i = \beta_0 + \beta^T x_i$

$\Rightarrow p_i = \sigma(z_i)$

$$\Rightarrow \frac{\partial L}{\partial \beta_0} = \frac{\partial}{\partial \beta_0}\left[-\sum_{i=1}^{N}\left\{y_i \ln(p_i) + (1-y_i)\ln(1-p_i)\right\}\right]$$

$$= \frac{\partial L}{\partial \beta_0}\left[-\sum_{i=1}^{N}\left\{y_i \ln(\sigma(z_i)) + (1-y_i)\ln(1-\sigma(z_i))\right\}\right]$$

$$= -\sum_{i=1}^{N}\left[y_i \frac{1}{p_i}\frac{\partial p_i}{\partial \beta_0} + (1-y_i)\frac{1}{1-p_i}\frac{\partial(1-p_i)}{\partial \beta_0}\right]$$

$$= -\sum_{i=1}^{N}\left[y_i \frac{p_i(1-p_i)}{p_i} + (1-y_i)\cdot\frac{-p_i(1-p_i)}{1-p_i}\right]$$

$$= -\sum_{i=1}^{N}\left[y_i(1-p_i) - (1-y_i)\cdot p_i\right]$$

$$= \sum\left[p_i - y_i p_i - y_i + y_i p_i\right]$$

$$= \sum\left[p_i - y_i\right]$$

$$\Rightarrow \frac{\partial L}{\partial \beta_0} = \sum_{i=1}^{N}(p_i - y_i)$$

$$\rightarrow \frac{\partial L}{\partial \beta} = \frac{\partial}{\partial \beta}\left[-\sum\left\{y_i \ln(p_i) + (1-y_i)\ln(1-p_i)\right\}\right]$$

$$\Rightarrow \frac{\partial L}{\partial \beta} = -\sum_{i=1}^{N}\left[y_i \cdot \frac{1}{p_i} \frac{\partial p_i}{\partial \beta} + (1-y_i) \cdot \frac{1}{1-p_i} \frac{\partial (1-p_i)}{\partial \beta} \right]$$

$$= -\sum_{i=1}^{N}\left[y_i \frac{p_i(1-p_i)x_i}{p_i} + (1-y_i) \cdot \frac{-p_i(1-p_i)x_i}{1-p_i} \right]$$

$$= -\sum_{i=1}^{N}\left[y_i(1-p_i)x_i - (1-y_i)p_i x_i \right]$$

$$= -\sum \left[x_i y_i - x_i y_i p_i - p_i x_i + x_i y_i p_i \right]$$

$$= -\sum \left[x_i y_i - p_i x_i \right]$$

$$\Rightarrow \frac{\partial L}{\partial \beta} = \sum_{i=1}^{N} (p_i - y_i) x_i$$

B.3 Addition of Regularization terms

→ The regularization or penalty, generally, is not applied on intercept term in order to allow flexibility for the bias term

→ Hence, for

(i) L_1 regularization:

Objective function: $-\sum_{i=1}^{N}\left[y_i \log(p_i) + (1-y_i)\log(1-p_i) \right] + \lambda \|\beta\|_1$

- L2 regularization:

Objective function:

$$-\sum_{i=1}^{N} [y_i \log(p_i) + (1-y_i)\log(1-p_i)] + \lambda \|\beta\|_2^2$$

→ If we do penalize the bias term as well, then the penalty term would look something like this:

- L_1 regularization: $\lambda \cdot \{|\beta_0| + |\beta_1|\}$
- L_2 regularization: $\lambda \cdot \{\beta_0^2 + \beta_1^2\}^{1/2}$

→ And the objective function would be the Negative Log likelihood linearly combined with the regularization term

i.e

- $L_1: -\sum [y_i \log(p_i) + (1-y_i)\log(1-p_i)] + \lambda(|\beta_0| + |\beta_1|)$
- $L_2: -\sum [y_i \log(p_i) + (1-y_i)\log(1-p_i)] + \lambda \cdot \{\beta_0^2 + \beta_1^2\}^{1/2}$

→ The effect of alteration on search space:

- On applying the L_1 regularization for given Logistic regression will alter the seach space in the shape of '<u>Diamond</u> or <u>Tetrahadron</u>'.
- While applying the L2 regularization, for given

logistic regression setup, it will alter the search space into a spherical or circular shape