Introduction to Statistical learning

Omid Safarzadeh

January 19, 2022

Table of contents

*Acknowledgement: This slide is prepared based on Murphy, 2012 and James et al., 2013

Logistic regression

- $Y \in \{0, 1\}$. Ex: $0 = ebola$, $1 =$ no ebola
- $X \in \mathscr{R}$

$$\pi(X) = \frac{e^{\beta_0+\beta_1 X}}{1 + e^{\beta_0+\beta_1 X}} \text{ (logistic function)}$$

- $\lim_{X\to-\infty} \pi(X)$? $\lim_{X\to+\infty} \pi(X)$?
- $\pi(X)$ models $Pr(Y = 1|X)$
- Odds:

$$\frac{\pi(X)}{1 - \pi(X)} = e^{\beta_0+\beta_1 X}$$

- Log-odds (logit):

$$logit(\pi(X)) = log(\frac{\pi(X)}{1 - \pi(X)}) = \beta_0 + \beta_1 X$$

- logit is linear in X!

MLE for simple logistic regression

- Data: $\mathscr{D} = \{(x_i, y_i)\}_{i=1}^{n}$
- Model: $Y_1, ..., Y_n$ are independent. $Y_i \sim$Bernoulli$(\pi(x_i))$

Likelihood

$$L(\beta_0, \beta_1) = p(\mathscr{D}|\beta_1, \beta_0) = \prod_{i:y_i=1} \pi(x_i) \prod_{i':y_{i'}=0} (1 - \pi(x_{i'}))$$

Log-likelihood

$$l(\beta_0, \beta_1) = \log p(\mathscr{D}|\beta_1, \beta_0) = \sum_{i=1}^{n} [y_i(\beta_0 + \beta_1 x_i) - log(1 + e^{\beta_0 + \beta_1 x_i})]$$

MLE

$$(\hat{\beta}_0^{MLE}, \hat{\beta}_1^{MLE}) = \arg_{\beta_0, \beta_1} \max L(\beta_0, \beta_1) = \arg_{\beta_0, \beta_1} \max l(\beta_0, \beta_1)$$

MLE for simple logistic regression

- No closed form solution for $(\hat{\beta}_0^{MLE}, \hat{\beta}_1^{MLE})$
- MLE can be found by **Newton-Raphson method**

Multiple logistic regression

- Response: $Y \in \{0, 1\}$
- Predictors: $\mathbf{X} = [1, X_1, ..., X_p]^T$
- Parameters: $\beta = [\beta_0, ..., \beta_p]^T$
- Logistic function:

$$\pi(\mathbf{X}; \beta) = \frac{e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p}} = \frac{e^{\beta^t \mathbf{X}}}{1 + e^{\beta^T \mathbf{X}}}$$

- $\pi(\mathbf{X}; \beta)$ models $Pr(Y = 1 | X_1, ..., X_p; \beta)$
- Odds:

$$\frac{\pi(\mathbf{X}; \beta)}{1 - \pi(\mathbf{X}; \beta)} = e^{\beta^T \mathbf{X}}$$

- Log-odds (logit):

$$logit(\pi(\mathbf{X}; \beta)) = log(\frac{\pi(\mathbf{X}; \beta)}{1 - \pi(\mathbf{X}; \beta)}) = \beta^T \mathbf{X}$$

- logit is linear in X!

MLE for multiple logistic regression

- Data: $\mathscr{D} = \{(\mathbf{x}_i, y_i)\}_{i=1}^n, \mathbf{x}_i = [1, x_{i1}, ..., x_{ip}]^T$
- Model: $Y_1, ..., Y_n$ are independent.

$$Y_i \sim \text{Bernoulli}(\pi(\mathbf{X}_i))$$

- Log-likelihood

$$l(\beta) = \log p(\mathscr{D}|\beta) = \sum_{i=1}^{n} [y_i \beta^T \mathbf{x}_i - log(1 + e^{\beta^T \mathbf{x}_i})]$$

- MLE

$$\hat{\beta}^{MLE} = \underset{\beta \in \mathscr{R}^{p+1}}{\text{arg}} \max l(\beta)$$

l_1 & l_2 regularized logistic regression

Optimization problem:

- l_1 regularized logistic regression:

$$(\hat{\beta}_0^{MLE}, \hat{\beta}^{MLE}) = \underset{\beta_0,\beta}{\arg\max}(\underbrace{\sum_i^n [y_i \beta^T \mathbf{x}_i - \log(1 + e^{\beta^T \mathbf{x}_i})] - \lambda \sum_{j=1}^p |\beta_j|)}_{\text{log-likelihood}}$$

- l_2 regularized logistic regression:

$$(\hat{\beta}_0^{MLE}, \hat{\beta}^{MLE}) = \underset{\beta_0,\beta}{\arg\max}(\underbrace{\sum_i^n [y_i \beta^T \mathbf{x}_i - \log(1 + e^{\beta^T \mathbf{x}_i})] - \lambda \sum_{j=1}^p \beta_j^2)}_{\text{log-likelihood}}$$

- What happens as $\lambda \to 0$?, $\lambda \to \infty$?

Logistic regression for $K > 2$ classes

- Response: $Y \in \mathscr{C}, \mathscr{C} = \{1, ..., K\}$.
- Predictors: $\mathbf{X} = [1, X_1, ..., X_p]^T$
- Parameters: $\beta_J = [\beta_{j0}, ..., \beta_{jp}]^T, j = 1, ..., K$

Model:

$$\log \frac{Pr(Y = 1|\mathbf{X} = \mathbf{x})}{Pr(Y = K|\mathbf{X} = \mathbf{x})} = \beta_1^T \mathbf{x}$$

$$\vdots$$

$$\log \frac{Pr(Y = 1|\mathbf{X} = \mathbf{x})}{Pr(Y = K|\mathbf{X} = \mathbf{x})} = \beta_{K-1}^T \mathbf{x}$$

- Parameters vector $\theta = \{\beta_1^T, ..., \beta_{K-1}^T\}$

Logistic regression for $K > 2$ classes

- Above equations can be solved for each

$$\pi_K(\mathbf{x};\theta) = Pr(Y = k|\mathbf{X} = \mathbf{x})$$

$$\pi_j(\mathbf{x};\theta) = \frac{exp(\beta_j^t \mathbf{x})}{1 + \sum_{l=1}^{K-1} exp(\beta_l^T \mathbf{x})}, \quad j = 1, ..., K-1$$

$$\pi_K(\mathbf{x};\theta) = \frac{1}{1 + \sum_{l=1}^{K-1} exp(\beta_l^T \mathbf{x})}$$

- Is $\pi_j(\mathbf{x};\theta) \in [0,1]$ for all $j = 1, ..., K$?
- What is $\sum_{j=1}^{K} \pi_j(\mathbf{x};\theta)$?

How to solve logistic regression for $K > 2$ classes

- Similar but not exactly the same as the case $K = 2$
- Apply Newton method.

References

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning: With applications in r*. Springer New York. https://books.google.fr/books?id=qcI%5C_AAAAQBAJ

Murphy, K. (2012). *Machine learning: A probabilistic perspective*. MIT Press. https://books.google.fr/books?id=NZP6AQAAQBAJ