

Ch 4.3.3 and 4.3.4 - Multiple and Multinomial Logistic Regression

Lecture 11 - CMSE 381

Prof. Mengsen Zhang

Michigan State University

::

Dept of Computational Mathematics, Science & Engineering

Fri, Sep 19, 2025

Announcements

CMSE381_F2025_Schedule : Schedule

Lec #	Date	Topic	Reading	HW
1	M 8/25	Intro / Python Review	1	
2	W 8/27	What is statistical learning	2.1	
3	F 8.29	Assessing Model Accuracy	2.2.1, 2.2.2	
	M 9/1	Labor Day - No Class		
4	W 9/3	Linear Regression	3.1	
5	F 9/5	More Linear Regression	3.1	HW #1 Due Sun 9/7
6	M 9/8	Multi-linear Regression	3.2	
7	W 9/10	Probably More Linear Regression	3.3	
8	F 9/12	Last of the Linear Regression		HW #2 Due Sun 9/14
9	M 9/15	Intro to classification, Bayes classifier, KNN classifier	2.2.3	
10	W 9/17	Logistic Regression	4.1, 4.2, 4.3.1-3	
11	F 9/19	Multiple Logistic Regression / Multinomial Logistic Regression	4.3.4-5	HW #3 Due Sun 9/21
	M 9/22	Project Day & Review		
	W 9/24	Midterm #1		
12	F 9/26	Leave one out CV	5.1.1, 5.1.2	

Announcements:

- Monday - Project day
 - Will talk about the project
- Monday - Review day
 - Send me your questions ([Link](#))
- Wednesday - Exam #1
 - Bring 8.5x11 sheet of paper
 - Handwritten both sides
 - Anything you want on it, but must be your work
 - You will turn it in
 - Bring Non-internet calculator

Last Time:

- Logistic Regression

This time:

- Multiple Logistic Regression
- Multinomial Logistic Regression

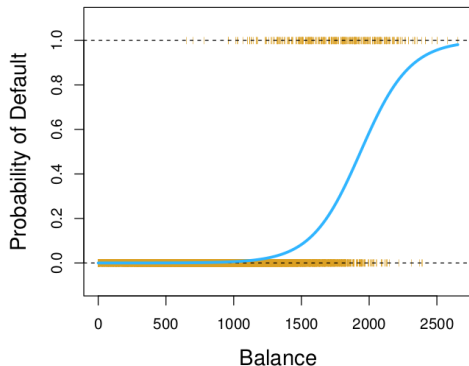
Section 1

Review of Logistic Regression from last time

Logistic regression

- Assume single input X
- Output takes values $Y \in \{\text{Yes}, \text{No}\}$

$$p(X) = \Pr(Y = \text{yes} \mid \text{balance})$$



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

How to get logistic function

Assume the (natural) log odds (logits) follow a linear model

$$\log \left(\frac{p(x)}{1 - p(x)} \right) = \beta_0 + \beta_1 x$$

Solve for $p(x)$:

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

Playing with the logistic function: desmos.com/calculator/cw1pyzzqci

Section 2

Multiple Logistic Regression

New assumption

$p \geq 1$ input variables

$$X_1, X_2, \dots, X_p$$

Y output variable has only two levels

Multiple features:

$$p(X) = \frac{e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}{1 + e^{\beta_0 + \beta_1 X_1 + \dots + \beta_p X_p}}$$

Equivalent to:

$$\log \left(\frac{p(X)}{1 - p(X)} \right) = \beta_0 + \beta_1 X_1 + \dots + \beta_p X_p$$

Example

	default	student	balance	income
0	No	No	729.526495	44361.625070
1	No	Yes	817.180407	12106.134700
2	No	No	1073.549164	31767.138950
3	No	No	529.250605	35704.493940
4	No	No	785.655883	38463.495880
5	No	Yes	919.588531	7491.558572
6	No	No	825.513331	24905.226580
7	No	Yes	808.667504	17600.451340
8	No	No	1161.057854	37468.529290
9	No	No	0.000000	29275.268290

Predict default from
balance, student, and income

Default data set

Section 3

Multinomial Logistic Regression

New assumption

$p \geq 1$ input variables

$$X_1, X_2, \dots, X_p$$

Y output variable has K levels

Remember dummy variables?

Slide from linear regression days

Region:

	x_{i1}	x_{i2}
South	1	0
West	0	1
East	0	0

Create spare dummy variables:

$$x_{i1} = \begin{cases} 1 & \text{if } i\text{th person from South} \\ 0 & \text{if } i\text{th person not from South} \end{cases}$$

$$x_{i2} = \begin{cases} 1 & \text{if } i\text{th person from West} \\ 0 & \text{if } i\text{th person not from West} \end{cases}$$

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \varepsilon_i$$

Example

Predict $Y \in \{\text{stroke}, \text{overdose}, \text{seizure}\}$ for hospital visits based on some input(s) X

$$\Pr(Y = \text{stroke} \mid X = x) =$$

$$\Pr(Y = \text{overdose} \mid X = x) =$$

$$\Pr(Y = \text{seizure} \mid X = x) =$$

Multinomial Logistic Regression

Plan A

- Assume Y has K levels
- Make K (the last one) the baseline

$$\Pr(Y = k|X = x) = \frac{e^{\beta_{k0} + \beta_{k1}x_1 + \cdots + \beta_{kp}x_p}}{1 + \sum_{l=1}^{K-1} e^{\beta_{l0} + \beta_{l1}x_1 + \cdots + \beta_{lp}x_p}}$$

$$\Pr(Y = K|X = x) = \frac{1}{1 + \sum_{l=1}^{K-1} e^{\beta_{l0} + \beta_{l1}x_1 + \cdots + \beta_{lp}x_p}}.$$

Example

Predict $Y \in \{\text{stroke}, \text{overdose}, \text{seizure}\}$ for hospital visits based on Xp

$$\Pr(Y = \text{stroke} \mid X = x) = \frac{\exp(\beta_{\text{str},0} + \beta_{\text{str},1}x)}{1 + \exp(\beta_{\text{str},0} + \beta_{\text{str},1}x) + \exp(\beta_{\text{OD},0} + \beta_{\text{OD},1}x)}$$

$$\Pr(Y = \text{overdose} \mid X = x) = \frac{\exp(\beta_{\text{OD},0} + \beta_{\text{OD},1}x)}{1 + \exp(\beta_{\text{str},0} + \beta_{\text{str},1}x) + \exp(\beta_{\text{OD},0} + \beta_{\text{OD},1}x)}$$

$$\Pr(Y = \text{seizure} \mid X = x) = \frac{1}{1 + \exp(\beta_{\text{str},0} + \beta_{\text{str},1}x) + \exp(\beta_{\text{OD},0} + \beta_{\text{OD},1}x)}$$

Log odds

Calculated so that log odds between *any pair of* classes is linear.
Specifically, for $Y = k$ vs $Y = K$, we have

$$\log \left(\frac{\Pr(Y = k \mid X = x)}{\Pr(Y = K \mid X = x)} \right) = \beta_{k0} + \beta_{k1}x_1 + \cdots + \beta_{kp}x_p$$

$$\Pr(Y = k \mid X = x) = \frac{e^{\beta_{k0} + \beta_{k1}x_1 + \cdots + \beta_{kp}x_p}}{1 + \sum_{l=1}^{K-1} e^{\beta_{l0} + \beta_{l1}x_1 + \cdots + \beta_{lp}x_p}}$$

$$\Pr(Y = K \mid X = x) = \frac{1}{1 + \sum_{l=1}^{K-1} e^{\beta_{l0} + \beta_{l1}x_1 + \cdots + \beta_{lp}x_p}}.$$

Plan B: Softmax coding

Treat all levels symmetrically

$$\Pr(Y = k|X = x) = \frac{e^{\beta_{k0} + \beta_{k1}x_1 + \cdots + \beta_{kp}x_p}}{\sum_{l=1}^K e^{\beta_{l0} + \beta_{l1}x_1 + \cdots + \beta_{lp}x_p}}.$$

Calculated so that log odds between two classes is linear

$$\log \left(\frac{\Pr(Y = k|X = x)}{\Pr(Y = k'|X = x)} \right) = (\beta_{k0} - \beta_{k'0}) + (\beta_{k1} - \beta_{k'1})x_1 + \cdots + (\beta_{kp} - \beta_{k'p})x_p.$$

Softmax example

$$\Pr(Y = \text{stroke} \mid X = x)$$

$$= \frac{\exp(\beta_{\text{str},0} + \beta_{\text{str},1}x)}{\exp(\beta_{\text{str},0} + \beta_{\text{str},1}x) + \exp(\beta_{\text{OD},0} + \beta_{\text{OD},1}x) + \exp(\beta_{\text{seiz},0} + \beta_{\text{seiz},1}x)}$$

$$\Pr(Y = \text{overdose} \mid X = x)$$

$$= \frac{\exp(\beta_{\text{OD},0} + \beta_{\text{OD},1}x)}{\exp(\beta_{\text{str},0} + \beta_{\text{str},1}x) + \exp(\beta_{\text{OD},0} + \beta_{\text{OD},1}x) + \exp(\beta_{\text{seiz},0} + \beta_{\text{seiz},1}x)}$$

$$\Pr(Y = \text{seizure} \mid X = x)$$

$$= \frac{\exp(\beta_{\text{seiz},0} + \beta_{\text{seiz},1}x)}{\exp(\beta_{\text{str},0} + \beta_{\text{str},1}x) + \exp(\beta_{\text{OD},0} + \beta_{\text{OD},1}x) + \exp(\beta_{\text{seiz},0} + \beta_{\text{seiz},1}x)}$$

Test your understanding: [PollEv](#)

Next time

CMSE381_F2025_Schedule : Schedule

Lec #		Date	Topic	Reading	HW
1	M	8/25	Intro / Python Review	1	
2	W	8/27	What is statistical learning	2.1	
3	F	8.29	Assessing Model Accuracy	2.2.1, 2.2.2	
	M	9/1	Labor Day - No Class		
4	W	9/3	Linear Regression	3.1	
5	F	9/5	More Linear Regression	3.1	HW #1 Due Sun 9/7
6	M	9/8	Multi-linear Regression	3.2	
7	W	9/10	Probably More Linear Regression	3.3	
8	F	9/12	Last of the Linear Regression		HW #2 Due Sun 9/14
9	M	9/15	Intro to classification, Bayes classifier, KNN classifier	2.2.3	
10	W	9/17	Logistic Regression	4.1, 4.2, 4.3.1-3	
11	F	9/19	Multiple Logistic Regression / Multinomial Logistic Regression	4.3.4-5	HW #3 Due Sun 9/21
	M	9/22	Project Day & Review		
	W	9/24	Midterm #1		
12	F	9/26	Leave one out CV	5.1.1, 5.1.2	