Ch 4.3.3 and 4.3.4 - Multiple and Multinomial Logistic Regression
Lecture 11 - CMSE 381

Prof. Elizabeth Munch

Michigan State University
::
Dept of Computational Mathematics, Science & Engineering

Fri, Sep 22, 2023

Announcements

Lec #	Date			Reading	Homeworks	*Quizzes* *(Note: These are not announced until after they happen)*
3	Fri	Sep 1	Assessing Model Accuracy	2.2.1, 2.2.2	HW #1 Due	Quiz #1
	Mon	Sep 4	No class - Labor day			
4	Wed	Sep 6	Linear Regression	3.1		
5	Fri	Sep 8	More Linear Regression	3.1/3.2		Quiz #2
6	Mon	Sep 11	Even more linear regression	3.2.2	Hw #2 Due	
7	Wed	Sep 13	Probably more linear regression	3.3		Quiz #3
8	Fri	Sep 15	Linear regression coding module			
9	Mon	Sep 18	Intro to classification, Bayes classifier, KNN classifier	2.2.3		
10	Wed	Sep 20	Logistic Regression	4.1, 4.2, 4.3.1-3		
11	Fri	Sep 22	Multiple Logistic Regression / Multinomial Logistic Regression /Project day	4.3.4-5	Hw #3 Due	
	Mon	Sep 25	*Review*			
	Wed	Sep 27	**Midterm #1**			
	Fri	Sep 29	No class - Dr Munch out of town			

Announcements:

- Homework #3 Due tonight on Crowdmark
- Monday - Review day
  - ▶ Nothing prepped
  - ▶ Bring your questions
- 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
  - ▶ Non-internet calculator if you want it

Covered in this lecture

Last Time:
- Logistic Regression

This time:
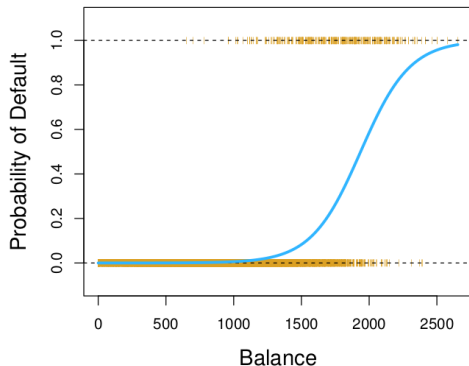- Multiple Logistic Regression
- Multinomial Logistic Regression

Section 1

Review of Logistic Regression from last time

Logistic regression

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

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

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

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

Y output variable has only two levels

$X_1, X_2, \cdots, X_p$
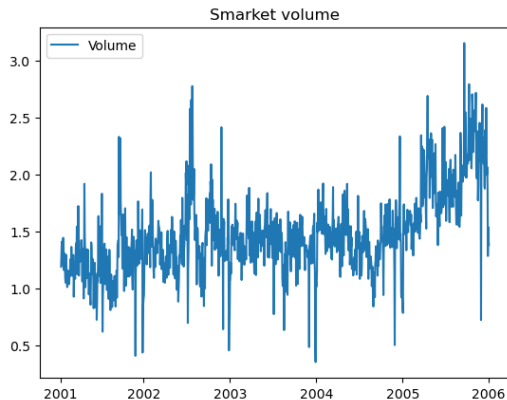
Multiple Logistic Regression

Multiple features:

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

Equivalent to:

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

Example from `Smarket` data

Smarket volume

	Lag1	Lag2	Volume	Direction
1	0.381	-0.192	1.19130	Up
2	0.959	0.381	1.29650	Up
3	1.032	0.959	1.41120	Down
4	-0.623	1.032	1.27600	Up
5	0.614	-0.623	1.20570	Up
...	...	...	...	...
1246	0.422	0.252	1.88850	Up
1247	0.043	0.422	1.28581	Down
1248	-0.955	0.043	1.54047	Up
1249	0.130	-0.955	1.42236	Down
1250	-0.298	0.130	1.38254	Down

1250 rows × 4 columns

Our Results

```python
X = smarket[['Lag1','Lag2','Volume']]
Y = smarket.Direction

clf = LogisticRegression(random_state=0)
clf.fit(X,Y)
```

```
         LogisticRegression
LogisticRegression(random_state=0)
```

$$p(X) = \frac{\exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)}{1 + \exp(\beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p)}$$

```python
print(clf.coef_)
print(clf.intercept_)
```

```
[[-0.07302967 -0.04272162  0.12862433]]
[-0.1158254]
```

$$p(X) = \frac{\exp(-0.115 - 0.073 \cdot \mathtt{Lag1} - 0.043 \cdot \mathtt{Lag2} + 0.129 \cdot \mathtt{Volume})}{1 + \exp(-0.115 - 0.073 \cdot \mathtt{Lag1} - 0.043 \cdot \mathtt{Lag2} + 0.129 \cdot \mathtt{Volume})}$$

Section 3

Multinomial Logistic Regression

$p \geq 1$ input variables

Y output variable has K levels

$X_1, X_2, \cdots, X_p$

Remember dummy variables?
Slide from linear regression days

Region:

	x_{i1}	x_{i2}
South		
West		
East		

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

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

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

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

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

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

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

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

Log odds

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

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

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

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

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

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

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

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

Section 4

Project day

Discussion and getting started

- Project description posted on github
- Today's goals:
  - ▶ Find one (or both) data sets
  - ▶ Mark which one you are using on the spreadsheet
  - ▶ Maybe get some data cleanup and exploration started

Next time

Lec #	Date			Reading	Homeworks	Quizzes *(Note: These are not announced until after they happen)*
3	Fri	Sep 1	Assessing Model Accuracy	2.2.1, 2.2.2	HW #1 Due	Quiz #1
	Mon	Sep 4	No class - Labor day			
4	Wed	Sep 6	Linear Regression	3.1		
5	Fri	Sep 8	More Linear Regression	3.1/3.2		Quiz #2
6	Mon	Sep 11	Even more linear regression	3.2.2	Hw #2 Due	
7	Wed	Sep 13	Probably more linear regression	3.3		Quiz #3
8	Fri	Sep 15	Linear regression coding module			
9	Mon	Sep 18	Intro to classification, Bayes classifier, KNN classifier	2.2.3		
10	Wed	Sep 20	Logistic Regression	4.1, 4.2, 4.3.1-3		
11	Fri	Sep 22	Multiple Logistic Regression / Multinomial Logistic Regression /Project day	4.3.4-5	Hw #3 Due	
	Mon	Sep 25	*Review*			
	Wed	Sep 27	**Midterm #1**			
	Fri	Sep 29	No class - Dr Munch out of town			