Ch 4.3 - Logistic Regression

Lecture 9 - CMSE 381

Prof. Elizabeth Munch

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

Weds, Sep 21, 2022

Announcements

Lec #		Date	Topic	Reading	Homeworks
1	W	Aug 31	Intro / First day stuff / Python Review Pt 1	1	
2	F	Sep 2	What is statistical learning?	2.1	
	M	Sep 5	No class - Labor day		
3	W	Sep 7	Assessing Model Accuracy	2.2.1, 2.2.2	HW #1 Due
4	F	Sep 9	Linear Regression	3.1	
5	M	Sep 12	More Linear Regression	3.1/3.2	
6	W	Sep 14	Even more linear regression	3.2.2	HW #2 Due
7	F	Sep 16	Probably more linear regression	3.3	
8	M	Sep 19	Intro to classification, Logisitic Regression	2.2.3, 4.1, 4.2, 4.3	
9	W	Sep 21	More logistic regression		HW #3 Due
10	F	Sep 23	Review		
11	M	Sep 26	***Midterm #1***		
12	W	Sep 28	[No class, Dr Munch out of town]		
13	F	Sep 30	[No class, Dr Munch out of town]		

Announcements:

- Homework #3 Due tonight on Crowdmark
- Friday - Review day
  - Nothing prepped
  - Bring your questions
- Monday - Exam #1
  - Bring 8.5x11 sheet of paper
  - Handwritten both sides
  - Ahything you want on it, but must be your work
  - You will turn it in

Covered in this lecture

Last Time:

- Classification basics
- KNN classifier
- Started Logistic Regression

This time:

- Logistic Regression

Section 1

Review of Classification and Logistic Regression from last time

Error rate

- Training data:
  $\{(x_1, y_1), \cdots, (x_n, y_n)\}$ with y_i qualitative
- Estimate $\hat{y} = \hat{f}(x)$
- Indicator variable

Training error rate:

$$\frac{1}{n}\sum_{i=1}^{n} \mathrm{I}(y_i \neq \hat{y}_i)$$

Test error rate:

$$\mathrm{Ave}(\mathrm{I}(y_0 \neq \hat{y}_0))$$

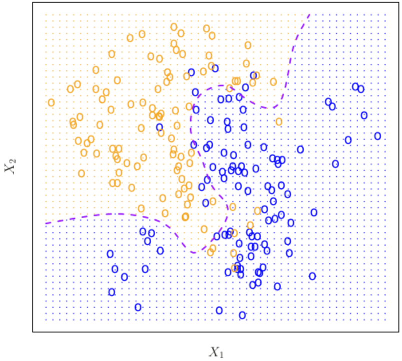
Best ever classifier

We can't have nice things

Bayes Classifier:
Give every observation the highest probability class given its predictor variables

$$\Pr(Y = j \mid X = x_0)$$

Bayes Decision Boundary



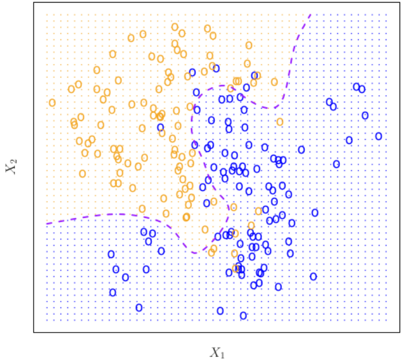
Bayes error rate

- Error at $X = x_0$

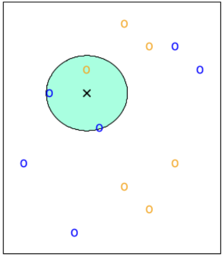
$$1 - \max_j \Pr(Y = j \mid X = x_0)$$

- Overall Bayes error:

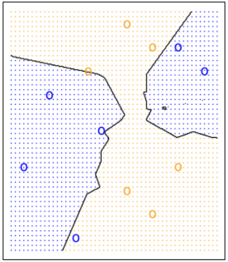
$$1 - E\left(\max_j \Pr(Y = j \mid X = x_0)\right)$$



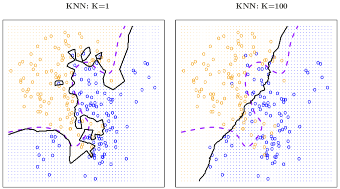
K-Nearest Neighbors



$K = 3$



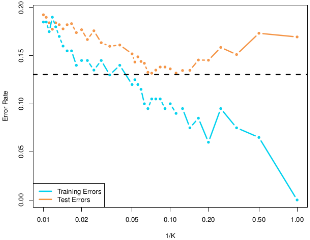
decision boundary



- Fix K positive integer
- $N(x)$ = the set of K closest neighbors to x
- Estimate conditional proability

$$\Pr(Y = j \mid X = x_0) = \frac{1}{K} \sum_{i \in N(x_0)} \mathrm{I}(y_i = j)$$

- Pick j with highest value



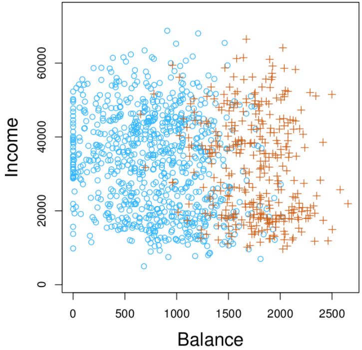
Section 2

Logistic Regression

What is classification

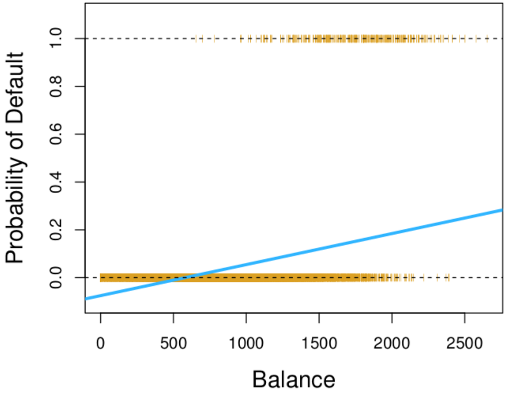
- Classification: When the response variable is qualitative
- Goal: Model the probability that Y belongs to a particular category

$$p(\texttt{balance}) = \Pr(\texttt{default} = \texttt{yes} \mid \texttt{balance})$$



Linear regression (Don't want to do this)

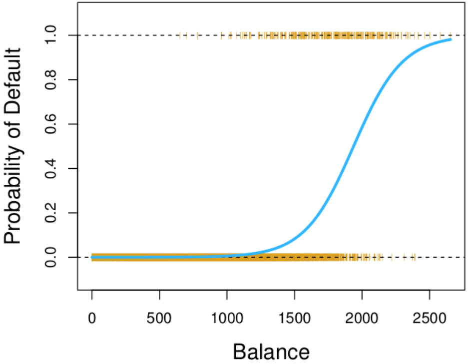
$$p(\texttt{balance}) = \Pr(\texttt{default} = \text{yes} \mid \texttt{balance})$$



$$p(\texttt{balance}) = \beta_0 + \beta_1 \texttt{balance}$$

Logistic regression (Do this)

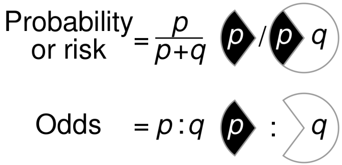
$$p(\texttt{balance}) = \Pr(\texttt{default} = \texttt{yes} \mid \texttt{balance})$$



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

Odds

$$\frac{p(x)}{1-p(x)} = \frac{\Pr(Y=1 \mid X=x)}{1-\Pr(Y=1 \mid X=x)} = \frac{\Pr(Y=1 \mid X=x)}{\Pr(Y=0 \mid X=x)}$$



Calculate the odds

If the probabilty of default is 90%, what are the odds?

If the odds are $1/3$, what is the probability of default?

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

Using coefficients to make predictions

	Coefficient	Std. error	z-statistic	p-value
Intercept	−10.6513	0.3612	−29.5	<0.0001
balance	0.0055	0.0002	24.9	<0.0001

What is the estimated probability of default for someone with a balance of \$1,000?

What is the estimated probability of default for someone with a balance of \$2,000:

$$\hat{p}(X) = \frac{e^{\hat{\beta}_0+\hat{\beta}_1 X}}{1+e^{\hat{\beta}_0+\hat{\beta}_1 X}} = \frac{e^{-10.6513+0.0055\times 2000}}{1+e^{-10.6513+0.0055\times 2000}} = 0.586$$

Interpreting the coefficients

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

	Coefficient	Std. error	z-statistic	p-value
`Intercept`	-10.6513	0.3612	-29.5	<0.0001
`balance`	0.0055	0.0002	24.9	<0.0001

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

Estimating Coefficients: Maximum Likelihood Estimation

- **Likelihood:** Probability that data is generated from a model

$$\ell(model) = \mathbb{P}[data \mid model]$$

- Find the most likely model

$$\max_{model} \ell(model)$$

- Hard to maximize likelihood, instead maximize log

$$\max_{model} \log(\ell(model))$$

- Strictly increasing log function doesn't change maximum

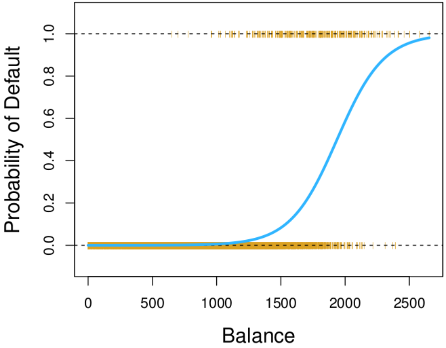
$$\Pr(Y = 1 \mid X) = p(X) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

$$\ell(\beta_0, \beta_1) = \prod_{i \mid y_i = 1} p(x_i) \prod_{i' \mid y_{i'} = 0} (1 - p(x_{i'}))$$

More on that likelihood function

$$\Pr(Y = 1 \mid X) = p(X)$$
$$= \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}}$$

$$\ell(\beta_0, \beta_1) = \prod_{i | y_i = 1} p(x_i) \prod_{i' | y_{i'} = 0} (1 - p(x_{i'}))$$



Confusion Matrix: Predicting `default` from `balance`

		True default status		
		No	Yes	Total
Predicted	No	9644	252	9896
default status	Yes	23	81	104
	Total	9667	333	10000

		True		
		Yes	No	Total
Predicted	Yes	a	b	$a+b$
	No	c	d	$c+d$
	Total	$a+c$	$b+d$	N

Do coding in jupyter notebook

Section 3

Multiple Logistic Regression

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:

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

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:

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

	Coefficient	Std. error	z-statistic	p-value
`Intercept`	-10.8690	0.4923	-22.08	<0.0001
`balance`	0.0057	0.0002	24.74	<0.0001
`income`	0.0030	0.0082	0.37	0.7115
`student[Yes]`	-0.6468	0.2362	-2.74	0.0062

Section 4

Multinomial Logistic Regression

Multinomial Logistic Regression

What if we have a categorical variable with more than two levels (let's say K of them)?

Plan A

Play the dummy variable game:

Make K the baseline:

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

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

Calculated so that log odds between two classes is linear:

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

Example

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

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

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

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

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

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

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

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

Next time

Lec #		Date	Topic	Reading	Homeworks
1	W	Aug 31	Intro / First day stuff / Python Review Pt 1	1	
2	F	Sep 2	What is statistical learning?	2.1	
	M	Sep 5	No class - Labor day		
3	W	Sep 7	Assessing Model Accuracy	2.2.1, 2.2.2	HW #1 Due
4	F	Sep 9	Linear Regression	3.1	
5	M	Sep 12	More Linear Regression	3.1/3.2	
6	W	Sep 14	Even more linear regression	3.2.2	HW #2 Due
7	F	Sep 16	Probably more linear regression	3.3	
8	M	Sep 19	Intro to classification, Logisitic Regression	2.2.3, 4.1, 4.2, 4.3	
9	W	Sep 21	More logistic regression		HW #3 Due
10	F	Sep 23	Review		
11	M	Sep 26	***Midterm #1***		
12	W	Sep 28	[No class, Dr Munch out of town]		
13	F	Sep 30	[No class, Dr Munch out of town]		