CS 66: Machine Learning

Prof. Sara Mathieson

Spring 2019

Admin

- Lab 3 due Thursday (you should be close to done by next lab)
- Midterm 1 Feb 27 (in lab)

Research Talk

Friday, February 15, 2019

11:30-12:30 p.m.

Science Center 240

Gerald Soosai Raj

University of Wisconsin – Madison

“Toward Making Computer Science Education Accessible and Relevant to Everyone”

Outline for February 15

- Logistic regression
- Likelihood functions
- Gradient descent for logistic regression

Outline for February 15

- Logistic regression
- Likelihood functions
- Gradient descent for logistic regression

Why is linear regression a bad choice for classification?

Case Study: you need to identify the medical condition of a patient in the emergency room on the basis of their symptoms.

Possible conditions (*y*) are:

- Stroke
- Drug overdoes
- Epileptic seizure

1) If you were forced to use linear regression for this problem, how could you encode *y* to make it real-valued?

2) What issues arise with making y real-valued?

3) What if you just had two outcomes (i.e. stroke and drug overdose) -- why is linear regression still not a good choice?

Why is linear regression a bad choice for classification?

Case Study: you need to identify the medical condition of a patient in the emergency room on the basis of their symptoms.

Possible conditions (y) are:

- Stroke
- Drug overdoes
- Epileptic seizure

1) If you were forced to use linear regression for this problem, how could you encode y to make it real-valued?

You could choose stroke=0, drug overdose=1, epileptic seizure=2 (or some permutation)

2) What issues arise with making y real-valued?

3) What if you just had two outcomes (i.e. stroke and drug overdose) -- why is linear regression still not a good choice?

Why is linear regression a bad choice for classification?

Case Study: you need to identify the medical condition of a patient in the emergency room on the basis of their symptoms.

Possible conditions (*y*) are:

- Stroke
- Drug overdoes
- Epileptic seizure

1) If you were forced to use linear regression for this problem, how could you encode *y* to make it real-valued?

You could choose stroke=0, drug overdose=1, epileptic seizure=2 (or some permutation)

2) What issues arise with making y real-valued?

Assumes some *ordering* of the outcomes that is probably not there!

3) What if you just had two outcomes (i.e. stroke and drug overdose) -- why is linear regression still not a good choice?

Why is linear regression a bad choice for classification?

Case Study: you need to identify the medical condition of a patient in the emergency room on the basis of their symptoms.

Possible conditions (*y*) are:

- Stroke
- Drug overdoes
- Epileptic seizure

1) If you were forced to use linear regression for this problem, how could you encode *y* to make it real-valued?

You could choose stroke=0, drug overdose=1, epileptic seizure=2 (or some permutation)

2) What issues arise with making y real-valued?

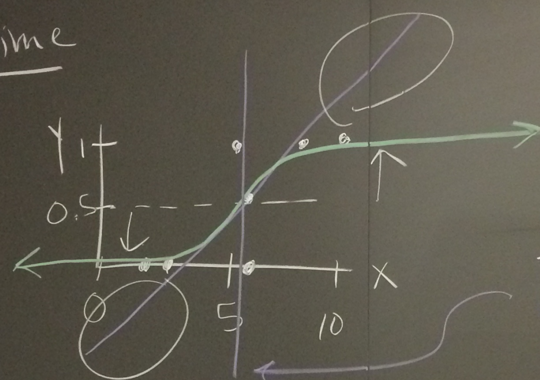
Assumes some *ordering* of the outcomes that is probably not there!

3) What if you just had two outcomes (i.e. stroke and drug overdose) -- why is linear regression still not a good choice?

The range of a linear function (i.e. *y* values) is $[-\infty, \infty]$, but we want $[0, 1]$

Last Time

x	y
8	1
6	0
2	0
5	1
10	1
3	0



$\vec{b}^T\vec{x} = 0$

linear regression:

$$[-\infty, \infty] \longrightarrow [-\infty, \infty]$$

want:

$$[-\infty, \infty] \longrightarrow [0, 1]$$

binary classification $\quad y \in \{0, 1\}$

$$h_{\vec{b}}(\vec{x}) = p(y=1 \mid \vec{x}, \vec{b})$$

$$h_{\vec{b}}(\vec{x}) = g(\vec{b}^T \vec{x})$$

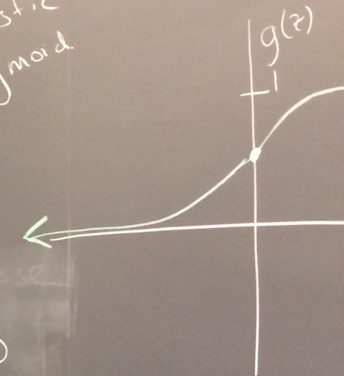
$\vec{b}^T \vec{x} \rightarrow$ linear model part $\rightarrow$ $g(z) = \frac{1}{1+e^{-z}}$ } logistic sigmoid

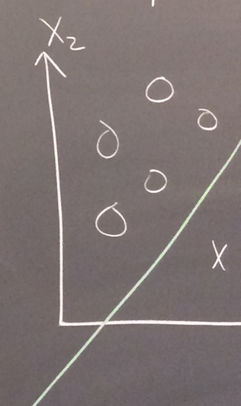
$z \to \infty, \; g(z) \to 1$

$z \to -\infty, \; g(z) \to 0$

$z = 0, \; g(z) = \frac{1}{2}$ ☆

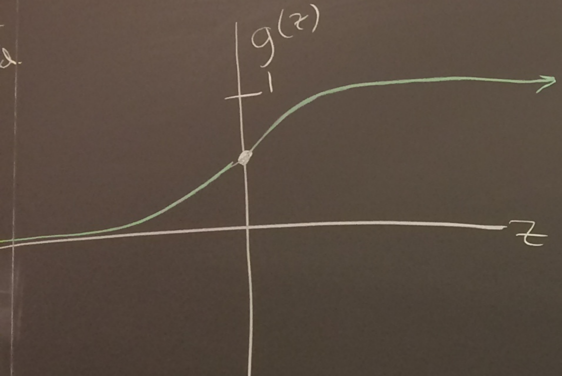
$$\boxed{\begin{array}{l} \vec{b}^T \vec{x} \geq 0 \Rightarrow y = 1 \\ \vec{b} \vec{x} < 0 \Rightarrow y = 0 \end{array}}$$

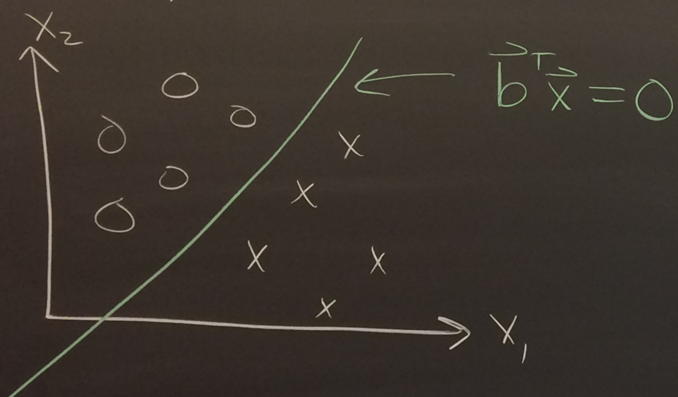




Outline for February 15

- Logistic regression
- Likelihood functions
- Gradient descent for logistic regression





Goal find $\vec{b}$.

- maximizing the likelihood of our data

Aside to likelihoods

Bernoulli Random Variable

n: coin flips $\quad H: 1$

p: prob of heads $\quad T: 0$

$\boxed{n=10}$: $\vec{y} = [0,0,1,1,0,1,0,1,0,0$

guess $\cancel{\boxed{p=0.001}}$

likeli

$L(p$

$\boxed{L(p}$

data

likelihood

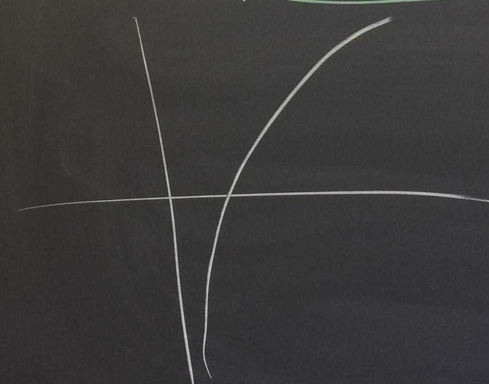
$$L(p) = \overbrace{(1-p)}^{\text{tails}}(1-p)\,p\,p \cdots\cdots (1-p)$$

$$= (1-p)^6 p^4$$

maximize!

$$\boxed{L(p) = \prod_{i=1}^{n} \underbrace{(p)^{y_i}}_{\substack{\text{prob} \\ y_i=1}} \underbrace{(1-p)}_{\substack{\text{prob} \\ y_i=0}}{}^{1-y_i}}$$

0, 1, 0, 1, 0, 0]



log

$\ell(p) =$

$\bar{Y} =$

So

log likelihood

$$\ell(p) = \log(L(p)) = \sum_{i=1}^{n} y_i \log p + \underbrace{\sum_{i=1}^{n} (1-y_i)}_{\sum 1 - \sum y_i} \log(1-p)$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

$$\log p \underbrace{\sum_{i=1}^{n} y_i}_{n\bar{y}}$$

$$f(x) = \log(x)$$

$$f'(x) = \frac{1}{x}$$

Solve for p

logistic Regression

$$L(\vec{b}) = \prod_{i=1}^{n} \underbrace{h_{\vec{b}}(\vec{x}_i)^{y_i}}_{\text{prob } y_i=1} \underbrace{\left(1-h_{\vec{b}}(\vec{x}_i)\right)^{1-y_i}}_{\text{prob } y_i=0}$$

$$g(z) = \frac{1}{1+e^{-z}}$$

$g'(z)$?

$$\ell(p) = n\bar{y}\log p + (n - n\bar{y})\log(1-p)$$

$$\frac{\partial \ell(p)}{\partial p} = \frac{\cancel{n}\bar{y}}{p} - \frac{\overset{1}{\cancel{n}} - \cancel{n}\bar{y}}{1-p} = 0$$

$$\frac{\bar{y}}{p} = \frac{1-\bar{y}}{1-p}$$

$$\Rightarrow \bar{y}(1-p) = p(1-\bar{y})$$

$$\bar{y} - \bar{y}p = p(1-\bar{y})$$

$$\bar{y} = p(1 \cancel{-\bar{y} + \bar{y}})$$

$$\Rightarrow \boxed{\hat{p} = \bar{y}}$$

maximum likelihood estimator for p

$$\boxed{\begin{gathered} \bar{y} = \frac{0\cdot 6 + 4\cdot 1}{10} \\ \hat{p} = \bar{y} = \frac{2}{5} = 0.4 \end{gathered}}$$