



# Logistic Regression

Data Science: Jordan Boyd-Graber  
University of Maryland

SLIDES ADAPTED FROM HINRICH SCHÜTZE

## What are we talking about?

- Statistical classification:  $p(y|x)$
- $y$  is typically a Bernoulli or multinomial outcome
- Classification uses: ad placement, spam detection
- Building block of other machine learning methods

## Logistic Regression: Definition

- Weight vector  $\beta_i$
- Observations  $X_i$
- “Bias”  $\beta_0$  (like intercept in linear regression)

$$P(Y = 0|X) = \frac{1}{1 + \exp[\beta_0 + \sum_i \beta_i X_i]} \quad (1)$$

$$P(Y = 1|X) = \frac{\exp[\beta_0 + \sum_i \beta_i X_i]}{1 + \exp[\beta_0 + \sum_i \beta_i X_i]} \quad (2)$$

- For shorthand, we'll say that

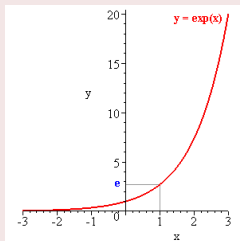
$$P(Y = 0|X) = \sigma(-(\beta_0 + \sum_i \beta_i X_i)) \quad (3)$$

$$P(Y = 1|X) = 1 - \sigma(-(\beta_0 + \sum_i \beta_i X_i)) \quad (4)$$

- Where  $\sigma(z) = \frac{1}{1 + \exp[-z]}$

## What's this “exp” doing?

### Exponential



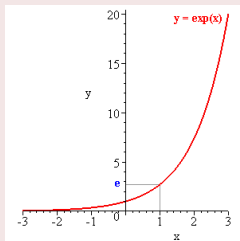
- $\exp[x]$  is shorthand for  $e^x$
- $e$  is a special number, about 2.71828
  - $e^x$  is the limit of compound interest formula as compounds become infinitely small
  - It's the function whose derivative is itself
- The “logistic” function is  $\sigma(z) = \frac{1}{1+e^{-z}}$
- Looks like an “S”
- Always between 0 and 1.

### Logistic



## What's this “exp” doing?

### Exponential



- $\exp[x]$  is shorthand for  $e^x$
- $e$  is a special number, about 2.71828
  - $e^x$  is the limit of compound interest formula as compounds become infinitely small
  - It's the function whose derivative is itself
- The “logistic” function is  $\sigma(z) = \frac{1}{1+e^{-z}}$
- Looks like an “S”
- Always between 0 and 1.
  - Allows us to model probabilities
  - Different from **linear** regression

### Logistic



## Logistic Regression Example

feature	coefficient	weight
bias	$\beta_0$	0.1
"viagra"	$\beta_1$	2.0
"mother"	$\beta_2$	-1.0
"work"	$\beta_3$	-0.5
"nigeria"	$\beta_4$	3.0

### Example 1: Empty Document?

$$X = \{\}$$

- What does  $Y = 1$  mean?

## Logistic Regression Example

feature	coefficient	weight
bias	$\beta_0$	0.1
"viagra"	$\beta_1$	2.0
"mother"	$\beta_2$	-1.0
"work"	$\beta_3$	-0.5
"nigeria"	$\beta_4$	3.0

- What does  $Y = 1$  mean?

### Example 1: Empty Document?

$$X = \{\}$$

- $P(Y = 0) = \frac{1}{1 + \exp[0.1]} =$
- $P(Y = 1) = \frac{\exp[0.1]}{1 + \exp[0.1]} =$

## Logistic Regression Example

feature	coefficient	weight
bias	$\beta_0$	0.1
"viagra"	$\beta_1$	2.0
"mother"	$\beta_2$	-1.0
"work"	$\beta_3$	-0.5
"nigeria"	$\beta_4$	3.0

- What does  $Y = 1$  mean?

### Example 1: Empty Document?

$$X = \{\}$$

- $P(Y = 0) = \frac{1}{1 + \exp[0.1]} = 0.48$
- $P(Y = 1) = \frac{\exp[0.1]}{1 + \exp[0.1]} = 0.52$
- Bias  $\beta_0$  encodes the prior probability of a class



## Logistic Regression Example

feature	coefficient	weight
bias	$\beta_0$	0.1
“viagra”	$\beta_1$	2.0
“mother”	$\beta_2$	-1.0
“work”	$\beta_3$	-0.5
“nigeria”	$\beta_4$	3.0

### Example 2

$X = \{\text{Mother, Nigeria}\}$

- What does  $Y = 1$  mean?

## Logistic Regression Example

feature	coefficient	weight
bias	$\beta_0$	0.1
"viagra"	$\beta_1$	2.0
"mother"	$\beta_2$	-1.0
"work"	$\beta_3$	-0.5
"nigeria"	$\beta_4$	3.0

- What does  $Y = 1$  mean?

### Example 2

$X = \{\text{Mother, Nigeria}\}$

- $P(Y = 0) = \frac{1}{1 + \exp[0.1 - 1.0 + 3.0]} =$
- $P(Y = 1) = \frac{\exp[0.1 - 1.0 + 3.0]}{1 + \exp[0.1 - 1.0 + 3.0]} =$
- Include bias, and sum the other weights

## Logistic Regression Example

feature	coefficient	weight
bias	$\beta_0$	0.1
"viagra"	$\beta_1$	2.0
"mother"	$\beta_2$	-1.0
"work"	$\beta_3$	-0.5
"nigeria"	$\beta_4$	3.0

- What does  $Y = 1$  mean?

### Example 2

$X = \{\text{Mother, Nigeria}\}$

- $P(Y = 0) = \frac{1}{1 + \exp[0.1 - 1.0 + 3.0]} = 0.11$
- $P(Y = 1) = \frac{\exp[0.1 - 1.0 + 3.0]}{1 + \exp[0.1 - 1.0 + 3.0]} = 0.88$
- Include bias, and sum the other weights

## Logistic Regression Example

feature	coefficient	weight
bias	$\beta_0$	0.1
“viagra”	$\beta_1$	2.0
“mother”	$\beta_2$	-1.0
“work”	$\beta_3$	-0.5
“nigeria”	$\beta_4$	3.0

### Example 3

$X = \{\text{Mother, Work, Viagra, Mother}\}$

- What does  $Y = 1$  mean?

## Logistic Regression Example

feature	coefficient	weight
bias	$\beta_0$	0.1
“viagra”	$\beta_1$	2.0
“mother”	$\beta_2$	-1.0
“work”	$\beta_3$	-0.5
“nigeria”	$\beta_4$	3.0

- What does  $Y = 1$  mean?

### Example 3

$X = \{\text{Mother, Work, Viagra, Mother}\}$

- $P(Y = 0) = \frac{1}{1 + \exp[0.1 - 1.0 - 0.5 + 2.0 - 1.0]} =$
- $P(Y = 1) = \frac{\exp[0.1 - 1.0 - 0.5 + 2.0 - 1.0]}{1 + \exp[0.1 - 1.0 - 0.5 + 2.0 - 1.0]} =$
- Multiply feature presence by weight

## Logistic Regression Example

feature	coefficient	weight
bias	$\beta_0$	0.1
“viagra”	$\beta_1$	2.0
“mother”	$\beta_2$	-1.0
“work”	$\beta_3$	-0.5
“nigeria”	$\beta_4$	3.0

- What does  $Y = 1$  mean?

### Example 3

$X = \{\text{Mother, Work, Viagra, Mother}\}$

- $P(Y = 0) = \frac{1}{1 + \exp[0.1 - 1.0 - 0.5 + 2.0 - 1.0]} = 0.60$
- $P(Y = 1) = \frac{\exp[0.1 - 1.0 - 0.5 + 2.0 - 1.0]}{1 + \exp[0.1 - 1.0 - 0.5 + 2.0 - 1.0]} = 0.30$
- Multiply feature presence by weight