Introduction to Machine Learning for Social Science

Class 4: Classification

Rochelle Terman

Postdoctoral Fellow
Center for International Security Cooperation
Stanford University

January 18th, 2018

Loose ends:
Confounding vs. collinearity


LIVE
SEN. HILLARY RODHAM CLINTON
D-New York
C-SPAN 2
U.S. Senate

Goal: predict Iraq vote (probability of yes, classify senators as for and against)
Method: Linear Probability Model & Logistic regression
Evaluation:

1) Accuracy
2) Precision
3) Recall

Key Terms:

- Classification
- Linear Probability Model
- Logit function and logit inverse function, Logistic regression
- Accuracy and the Precision/Recall Tradeoff

Key Techniques and `R` Functions

- `glm`
- Natural logarithm `log`
- `subset` , `cbind`
- `table`

To the R Code! (Section 1)

Classification

- Vote: Yay / Nay?

Classification

- Vote: Yay / Nay?
- Email: Spam / Not Spam?

Classification

- Vote: Yay / Nay?
- Email: Spam / Not Spam?
- Online transaction: Faudulent (Yes / No)?

Classification

- Vote: Yay / Nay?
- Email: Spam / Not Spam?
- Online transaction: Faudulent (Yes / No)?

$$y \in \{0, 1\} \qquad \begin{bmatrix} 0\text{: "Negative class"} \\ 1\text{: "Positive class"} \end{bmatrix}$$

Two Estimation Goals

Estimate:

Two Estimation Goals

Estimate:

- Probability of voting yes: $\Pr(\widehat{\text{Vote}_i = 1}|\boldsymbol{x}_i)$

Two Estimation Goals

Estimate:

- Probability of voting yes: $\Pr(\widehat{\text{Vote}_i = 1}|\boldsymbol{x}_i)$
- Classification of vote: $\widehat{\text{Vote}_i} = I(\Pr(\widehat{\text{Vote}_i = 1}|\boldsymbol{x}_i) > t)$, where t is a threshold

Linear Probability Model

Linear Probability Model

$$\text{Vote}_i = \boldsymbol{\beta} \cdot \boldsymbol{x}_i + \epsilon_i$$

Linear Probability Model

$$\begin{aligned} \text{Vote}_i &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i + \epsilon_i \\ \widehat{\Pr(\text{Vote}_i = 1 | \boldsymbol{X}_i)} &= \widehat{\boldsymbol{\beta}} \cdot \boldsymbol{x}_i \end{aligned}$$

Linear Probability Model

$$
\begin{aligned}
\text{Vote}_i &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i + \epsilon_i \\
\widehat{\Pr(\text{Vote}_i = 1 | \boldsymbol{X}_i)} &= \widehat{\boldsymbol{\beta}} \cdot \boldsymbol{x}_i \\
\widehat{\text{Vote}_i} &= 1 \text{ if } \widehat{\boldsymbol{\beta}} \cdot \boldsymbol{x}_i > t
\end{aligned}
$$

Linear Probability Model

$$
\begin{aligned}
\text{Vote}_i &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i + \epsilon_i \\
\widehat{\Pr(\text{Vote}_i = 1 | \boldsymbol{X}_i)} &= \widehat{\boldsymbol{\beta}} \cdot \boldsymbol{x}_i \\
\widehat{\text{Vote}}_i &= 1 \text{ if } \widehat{\boldsymbol{\beta}} \cdot \boldsymbol{x}_i > t \\
\widehat{\text{Vote}}_i &= 0 \text{ if } \widehat{\boldsymbol{\beta}} \cdot \boldsymbol{x}_i \leq t
\end{aligned}
$$

Linear Probability Model

$$
\begin{aligned}
\text{Vote}_i &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i + \epsilon_i \\
\widehat{\Pr(\text{Vote}_i = 1 | \boldsymbol{X}_i)} &= \widehat{\boldsymbol{\beta}} \cdot \boldsymbol{x}_i \\
\widehat{\text{Vote}}_i &= 1 \text{ if } \widehat{\boldsymbol{\beta}} \cdot \boldsymbol{x}_i > t \\
\widehat{\text{Vote}}_i &= 0 \text{ if } \widehat{\boldsymbol{\beta}} \cdot \boldsymbol{x}_i \leq t
\end{aligned}
$$

`R Code (Section 2)`

(Potential) Problems with Linear Probability Model

- Probabilities greater than 1, less than 0
- Potentially implausible relationship between covariates and response

(Potential) Problems with Linear Probability Model

- Probabilities greater than 1, less than 0
- Potentially implausible relationship between covariates and response

Solution: Logistic Regression: $0 \leq f(X) \leq 1$

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.
- We'll call $\log_e$ natural logarithm. And we'll assume $\log = \log_e$

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.
- We'll call $\log_e$ natural logarithm. And we'll assume $\log = \log_e$
- $\log e = 1$ (because $e^1 = e$)

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.
- We'll call $\log_e$ natural logarithm. And we'll assume $\log = \log_e$
- $\log e = 1$ (because $e^1 = e$)
- $\log_{10} 1000 = 3$ (because $10^3 = 1000$)

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.
- We'll call $\log_e$ natural logarithm. And we'll assume $\log = \log_e$
- $\log e = 1$ (because $e^1 = e$)
- $\log_{10} 1000 = 3$ (because $10^3 = 1000$)

Some rules of logarithms

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.
- We'll call $\log_e$ natural logarithm. And we'll assume $\log = \log_e$
- $\log e = 1$ (because $e^1 = e$)
- $\log_{10} 1000 = 3$ (because $10^3 = 1000$)

Some rules of logarithms

- $\exp(\log(a)) = e^{\log(a)} = a$

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.
- We'll call $\log_e$ natural logarithm. And we'll assume $\log = \log_e$
- $\log e = 1$ (because $e^1 = e$)
- $\log_{10} 1000 = 3$ (because $10^3 = 1000$)

Some rules of logarithms

- $\exp(\log(a)) = e^{\log(a)} = a$
- $\log(a \times b) = \log(a) + \log(b)$ (!!!!!!)

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.
- We'll call $\log_e$ natural logarithm. And we'll assume $\log = \log_e$
- $\log e = 1$ (because $e^1 = e$)
- $\log_{10} 1000 = 3$ (because $10^3 = 1000$)

Some rules of logarithms

- $\exp(\log(a)) = e^{\log(a)} = a$
- $\log(a \times b) = \log(a) + \log(b)$ (!!!!!!)
- $\log(\frac{a}{b}) = \log(a) - \log(b)$

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.
- We'll call $\log_e$ natural logarithm. And we'll assume $\log = \log_e$
- $\log e = 1$ (because $e^1 = e$)
- $\log_{10} 1000 = 3$ (because $10^3 = 1000$)

Some rules of logarithms

- $\exp(\log(a)) = e^{\log(a)} = a$
- $\log(a \times b) = \log(a) + \log(b)$ (!!!!!!)
- $\log(\frac{a}{b}) = \log(a) - \log(b)$
- $\log(a^b) = b \log(a)$

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.
- We'll call $\log_e$ natural logarithm. And we'll assume $\log = \log_e$
- $\log e = 1$ (because $e^1 = e$)
- $\log_{10} 1000 = 3$ (because $10^3 = 1000$)

Some rules of logarithms

- $\exp(\log(a)) = e^{\log(a)} = a$
- $\log(a \times b) = \log(a) + \log(b)$ (!!!!!!)
- $\log(\frac{a}{b}) = \log(a) - \log(b)$
- $\log(a^b) = b \log(a)$
- $\log(1) = 0$

A Brief Reminder About (Natural) Logarithms

Logarithm log is a class of functions.

- $\log_e z = x$ if $e^x = z$.
- We'll call $\log_e$ natural logarithm. And we'll assume $\log = \log_e$
- $\log e = 1$ (because $e^1 = e$)
- $\log_{10} 1000 = 3$ (because $10^3 = 1000$)

Some rules of logarithms

- $\exp(\log(a)) = e^{\log(a)} = a$
- $\log(a \times b) = \log(a) + \log(b)$ (!!!!!!)
- $\log(\frac{a}{b}) = \log(a) - \log(b)$
- $\log(a^b) = b \log(a)$
- $\log(1) = 0$

Call $p_i = \Pr(\text{Vote}_i = 1|\boldsymbol{x}_i)$

Call $p_i = \Pr(\text{Vote}_i = 1|\boldsymbol{x}_i)$

$$\text{Vote}_i \sim \text{Bernoulli}(p_i)$$

Call $p_i = \Pr(\text{Vote}_i = 1|\boldsymbol{x}_i)$

$$\begin{aligned} \text{Vote}_i &\sim \text{Bernoulli}(p_i) \\ p_i &= f(\boldsymbol{\beta} \cdot \boldsymbol{x}_i) \end{aligned}$$

Call $p_i = \Pr(\text{Vote}_i = 1|\boldsymbol{x}_i)$

$$\begin{aligned} \text{Vote}_i &\sim \text{Bernoulli}(p_i) \\ p_i &= f(\boldsymbol{\beta} \cdot \boldsymbol{x}_i) \\ \log\left(\frac{p_i}{1-p_i}\right) &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i \end{aligned}$$

Call $p_i = \Pr(\text{Vote}_i = 1|\boldsymbol{x}_i)$

$$
\begin{aligned}
\text{Vote}_i &\sim \text{Bernoulli}(p_i) \\
p_i &= f(\boldsymbol{\beta} \cdot \boldsymbol{x}_i) \\
\log\left(\frac{p_i}{1-p_i}\right) &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i \\
p_i &= \frac{\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}
\end{aligned}
$$

Call $p_i = \Pr(\text{Vote}_i = 1|\boldsymbol{x}_i)$

$$
\begin{aligned}
\text{Vote}_i &\sim \text{Bernoulli}(p_i) \\
p_i &= f(\boldsymbol{\beta} \cdot \boldsymbol{x}_i) \\
\log\left(\frac{p_i}{1-p_i}\right) &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i \\
p_i &= \frac{\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)} \\
&= \frac{1}{1+\exp(-\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}
\end{aligned}
$$

Call $p_i = \Pr(\text{Vote}_i = 1|\boldsymbol{x}_i)$

$$\begin{aligned}
\text{Vote}_i &\sim \text{Bernoulli}(p_i) \\
p_i &= f(\boldsymbol{\beta} \cdot \boldsymbol{x}_i) \\
\log\left(\frac{p_i}{1-p_i}\right) &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i \\
p_i &= \frac{\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)} \\
&= \frac{1}{1+\exp(-\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}
\end{aligned}$$

Important functions:

Call $p_i = \Pr(\text{Vote}_i = 1|\boldsymbol{x}_i)$

$$\begin{aligned}
\text{Vote}_i &\sim \text{Bernoulli}(p_i) \\
p_i &= f(\boldsymbol{\beta} \cdot \boldsymbol{x}_i) \\
\log\left(\frac{p_i}{1-p_i}\right) &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i \\
p_i &= \frac{\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)} \\
&= \frac{1}{1+\exp(-\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}
\end{aligned}$$

Important functions:

$$\text{odds}(p) = \frac{p}{1-p}$$

Call $p_i = \Pr(\text{Vote}_i = 1|\boldsymbol{x}_i)$

$$\begin{aligned}
\text{Vote}_i &\sim \text{Bernoulli}(p_i) \\
p_i &= f(\boldsymbol{\beta} \cdot \boldsymbol{x}_i) \\
\log\left(\frac{p_i}{1-p_i}\right) &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i \\
p_i &= \frac{\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)} \\
&= \frac{1}{1+\exp(-\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}
\end{aligned}$$

Important functions:

$$\begin{aligned}
\text{odds}(p) &= \frac{p}{1-p} \\
\text{log odds or logit}(p) &= \log\left(\frac{p}{1-p}\right)
\end{aligned}$$

Call $p_i = \Pr(\text{Vote}_i = 1|\boldsymbol{x}_i)$

$$\begin{aligned}
\text{Vote}_i &\sim \text{Bernoulli}(p_i) \\
p_i &= f(\boldsymbol{\beta} \cdot \boldsymbol{x}_i) \\
\log\left(\frac{p_i}{1-p_i}\right) &= \boldsymbol{\beta} \cdot \boldsymbol{x}_i \\
p_i &= \frac{\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}{1+\exp(\boldsymbol{\beta} \cdot \boldsymbol{x}_i)} \\
&= \frac{1}{1+\exp(-\boldsymbol{\beta} \cdot \boldsymbol{x}_i)}
\end{aligned}$$

Important functions:

$$\begin{aligned}
\text{odds}(p) &= \frac{p}{1-p} \\
\text{log odds or logit}(p) &= \log\left(\frac{p}{1-p}\right) \\
\text{logistic function or logit}^{-1}(a) &= \frac{1}{1+\exp(-a)}
\end{aligned}$$

R Code (Section 3)

How do we interpret β?

Linear Regression: $\beta_1 \rightsquigarrow$ average change in Y with one-unit increase in X_1.

How do we interpret β?

Linear Regression: $\beta_1 \rightsquigarrow$ average change in Y with one-unit increase in X_1.

Logistic Regression: $\beta_1 \rightsquigarrow$ average change in *log odds* with one-unit increase in X_1

How do we interpret β?

Linear Regression: $\beta_1 \rightsquigarrow$ average change in Y with one-unit increase in X_1.

Logistic Regression: $\beta_1 \rightsquigarrow$ average change in *log odds* with one-unit increase in X_1

- Non-linear relationship between X and $p(X)$.

How do we interpret β?

Linear Regression: $\beta_1 \rightsquigarrow$ average change in Y with one-unit increase in X_1.

Logistic Regression: $\beta_1 \rightsquigarrow$ average change in *log odds* with one-unit increase in X_1

- Non-linear relationship between X and $p(X)$.
- Amount that $p(X)$ changes due to one-unit change in X will depend on current value of X.

How do we interpret β?

Linear Regression: $\beta_1 \rightsquigarrow$ average change in Y with one-unit increase in X_1.

Logistic Regression: $\beta_1 \rightsquigarrow$ average change in *log odds* with one-unit increase in X_1

- Non-linear relationship between X and $p(X)$.
- Amount that $p(X)$ changes due to one-unit change in X will depend on current value of X.
- Regardless of value of X, if $\beta_1 > 0$, then increasing $X \rightsquigarrow$ increasing $p(X)$.

How do we interpret β?

Linear Regression: $\beta_1 \rightsquigarrow$ average change in Y with one-unit increase in X_1.

Logistic Regression: $\beta_1 \rightsquigarrow$ average change in *log odds* with one-unit increase in X_1

- Non-linear relationship between X and $p(X)$.
- Amount that $p(X)$ changes due to one-unit change in X will depend on current value of X.
- Regardless of value of X, if $\beta_1 > 0$, then increasing $X \rightsquigarrow$ increasing $p(X)$.
- odds ratio: e^{β_1}, represents how the *odds* change with a 1 unit increase in β_1 holding all other variables constant. Remains constant for any value of X.

Up next:
Fitting the Logistic Regression