Foundations of Data Science: Logistic regression
Overview

- Principle of Logistic Regression
- Interpretation of Logistic Regression coefficients
- Multiple Logistic Regression
- Logistic Regression as a classifier
- (Maximum likelihood estimation of Logistic regression coefficients)
Foundations of Data Science:
Logistic regression –
Principle of logistic regression
Supervised classification

Binary (or dichotomous) response variable: Credit

Approved
Not approved
Classification task on one continuous variable
Exercise:

(a) Draw a linear regression line through this data.

(b) Convert the linear regression prediction to a predicted class label (0/1).

(c) Where is the decision boundary?

(d) How many classification errors are more?
(a) Draw Linear regression line through this toy data.

(b) Convert the linear regression prediction to a predicted class label (0/1).

(c) Where is the decision boundary?

(d) How many classification errors are there?
Logistic function

\[ f(x) = \frac{e^x}{1 + e^x} = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}} \]

\[ P(Y = 1 \mid x = x) = f(\hat{\beta}_0 + \hat{\beta}_1 x) = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}} \]

\[ \beta_1 \text{ controls slope} \]

\[ \beta_0 \text{?} \]
Application to continuous variable in credit example

\[ \hat{\beta}_0 = -1.176 \]

\[ \hat{\beta}_1 = 0.031 \]
**Binary variables: odds and odds ratios**

\[
P(y = y \mid x = x)
\]

<table>
<thead>
<tr>
<th>Employed</th>
<th>Approved</th>
<th>Not approved</th>
<th>Approval odds</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.25</td>
<td>0.75</td>
<td>0.34</td>
</tr>
<tr>
<td>1</td>
<td>0.71</td>
<td>0.29</td>
<td>2.42</td>
</tr>
</tbody>
</table>

\[
y \in \{ "Not approved", "Approved" \}
\]

\[
x \in \{ "Not Emp.", "Emp." \}
\]

**Odds (Success) =** \[
\frac{P(\text{Success})}{P(\text{Failure})}
\]

**Odds ratio OR (x) =** \[
\frac{\text{Odds (Success) } | x = \text{True}}{\text{Odds (Success) } | x = \text{False}}
\]

\[
\text{App.} \quad 0.71 \quad 0.29 \quad 2.42
\]

\[
\text{OR (x)} = \frac{2.42}{0.34} = 7.09
\]

**Effect size C**

\[
\text{C} = 6.09 \%
\]
Foundations of Data Science:
Logistic regression –
Interpretation of logistic regression coefficients
Interpretation of $\hat{\beta}_0$

**Linear Regression (Lin reg):**

- $y$-axis
- $x$-axis
- $\beta_0$ and $\beta_1$

**Logistic Regression (Log. reg):**

- $y$-axis
- $x$-axis
- $f(\hat{\beta}_0) = P(\text{success} \mid x=0)$
- $f(\hat{\beta}_0 + 0 \cdot \hat{\beta}_1 ) = f(\hat{\beta}_0) = \frac{1}{1 + e^{-\hat{\beta}_0}}$
\[ f(\hat{\beta}_0) = f(-1.176) = 0.236 \]

\[ \hat{\beta}_0 = -1.176 \]
\[ \hat{\beta}_1 = 0.031 \]
Log odds

\[
\log \text{Odds}(\text{success}) = \ln \frac{P(\text{success})}{P(\text{failure})}
\]

\[
= \ln \frac{P(\text{success})}{1 - P(\text{success})}
\]

\[
\Rightarrow \log \text{odds} + 1 = \text{odds} \times e
\]

<table>
<thead>
<tr>
<th>\text{p}</th>
<th>\text{odds}</th>
<th>\log \text{odds}</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.5</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>&gt;0.5</td>
<td>&gt;1</td>
<td>&gt;0</td>
</tr>
<tr>
<td>&lt;0.5</td>
<td>&lt;1</td>
<td>&lt;0</td>
</tr>
<tr>
<td>1</td>
<td>→ ∞</td>
<td>→ ∞</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>→ -∞</td>
</tr>
</tbody>
</table>
Log odds $+1 \Rightarrow$ odds increase by factor $e$

logistic unit = logit

$\hat{\beta}_o = -1.176$ logits

$\text{logit } (p) = \ln \frac{p}{1-p}$
Logistic Regression in terms of log odds

1. Success: \( P(Y=1|x) = f(\beta_0 + \beta_1 x) = \frac{1}{1 + e^{-\beta_0 - \beta_1 x}} \)

2. Failure: \( P(Y=0|x) = 1 - f(\beta_0 + \beta_1 x) = 1 - \frac{1}{1 + e^{-\beta_0 - \beta_1 x}} \)

\[
\frac{e^{\beta_0 - \beta_1 x}}{1 + e^{-\beta_0 - \beta_1 x}}
\]

Odds: \( \frac{P(Y=1|x)}{P(Y=0|x)} = \frac{1}{e^{-\beta_0 - \beta_1 x}} = e^{\beta_0 + \beta_1 x} \)

Log odds: \( \ln \frac{P(Y=1|x)}{P(Y=0|x)} = \beta_0 + \beta_1 x = \text{logit} (P(Y=1|x)) \)
Interpretation of $\hat{\beta}$.

\[
\text{Odds}(x) = e^{\hat{\beta}_0 + \hat{\beta}_1 x}
\]

\[
= e^{\hat{\beta}_0} e^{\hat{\beta}_1 x}
\]

$x = \{0, 1\}$  \quad \text{OR}(x) = \frac{\text{Odds}(1)}{\text{Odds}(0)}

\[
\text{OR}(x) = e^{\hat{\beta}_1}
\]

\[
\log \text{OR}(x) = \hat{\beta}_1.
\]

(credit e.g.  \quad \text{OR}(\text{Age}) = e^{0.03} \approx 1.03$
Foundations of Data Science:
Logistic regression –
Multiple logistic regression
Principle of multiple logistic regression

Predictor variables

\( x^{(1)} : \text{Age} \)

\( x^{(2)} : \text{Employment} = \{0, 1\} \)

\[
\Pr(Y = 1 \mid x^{(1)}, x^{(2)}, \ldots) = \sigma (\beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \ldots)
\]
Multiple logistic regression applied to the credit example

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>Odds or OR</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\beta}_0$</td>
<td>Intercept</td>
<td>-1.969</td>
</tr>
<tr>
<td>$\hat{\beta}_1$</td>
<td>Age</td>
<td>0.029</td>
</tr>
<tr>
<td>$\hat{\beta}_2$</td>
<td>Employed</td>
<td>1.881</td>
</tr>
</tbody>
</table>

\[ \text{odds} = e^{\beta} \]
\[ \text{OR} = e^{\beta} \]
\[ \log \text{odds} = \log \text{it's} \]
\[ e^{\beta} \]
Bootstrap confidence intervals

Does age affect credit approval?

\[ H_0: \text{age does not affect credit approval } \Rightarrow e^{\hat{\beta}_1} = 1 \]

\[ H_a: \text{“” affect credit approval in some way.} \]
Discussion question

Can you think of any problems in the reasoning that we’ve used to suggest age and credit approval are related?
Approval

Employment

Credit score

Age

Lurking variable

Mediator variable
Foundations of Data Science:
Logistic regression – The logistic regression regression classifier
Converting logistic regression to a classifier

- Fit logistic regression model
- Set threshold $c$ in terms of log odds and apply to predicted log odds:

\[
\hat{\beta}_0 + \hat{\beta}_1 x^{(1)} + \hat{\beta}_2 x^{(2)} + \ldots \geq c \Rightarrow \hat{y} = 1
\]
\[
\hat{\beta}_0 + \hat{\beta}_1 x^{(1)} + \hat{\beta}_2 x^{(2)} + \ldots < c \Rightarrow \hat{y} = 0
\]

\[c = 0 \Rightarrow \text{odds of } 1 = p = 0.5\]
Decision boundary

\[ C = \ln \frac{1}{3} = -\ln 3 \]

\[ \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \ldots > c \Rightarrow \hat{y} = 1 \]

\[ \beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} + \ldots \leq c \Rightarrow \hat{y} = 0 \]
Ethics: logistic regression can be transparent

Credit scoring system:

• If you are in employment you score 1.625, if not you score 0

• Multiply your age by 0.029 and add the result to your score

• Round your income to the nearest 1000. Multiply the number of zeros in this figure by 0.320 and add the result to your score

• If you scored more than 2.246, your credit will be approved

Cf. "Promote Values of Transparency, Autonomy and Trustworthiness" (Vallor, 2018)
Logistic regression versus k-NN

Decision boundary, flexibility/over-fitting, transparency

Standardised input variables
Summary

- Interpret $\hat{\beta}_0$ and $\hat{\beta}_1$ in terms of log odds

- Extend logistic regression to multiple variables

- Use logistic regression as a classifier

- Not covered (yet): derivation of logistic regression from principle of max likelihood
Foundations of Data Science: Logistic regression – Maximum likelihood estimation of logistic regression coefficients
Principle of Maximum Likelihood

\[ y = \beta_0 + \beta_1 x \]

\[ \hat{y}_i = \beta_0 + \beta_1 x \]

\[ \sum_{i=1}^{n} (y_i - \beta_0 + \beta_1 x_i)^2 \]

Optimise \( \beta_0, \beta_1 \):

\[ \hat{\beta}_0 = \hat{\beta}_1 = \ldots \]

Adjust coefficients so as to maximise the likelihood of the data.

\( \Rightarrow \) Expression for max. likelihood

Optimise w.r.t. \( \beta_0, \beta_1, \ldots \).
Likelihood of one point

\[ P (Y_i = 1 \mid X_i = x_i) = f (\beta_0 + \beta_1 x_i) \]  
\[ P (Y_i = 0 \mid X_i = x_i) = 1 - f (\beta_0 + \beta_1 x_i) \]

\[ \Rightarrow P (Y_i = y_i \mid X_i = x_i) = 
\begin{align*}
  y_i & \cdot f (\beta_0 + \beta_1 x_i) + (1 - y_i) (1 - f (\beta_0 + \beta_1 x_i)) 
\end{align*} \]
Likelihood of data given model

Assumption: responses are independent, given value of predictor variables

\[
P(Y = y \mid X = x) = P(Y = y_1 \mid X = x_1) P(Y = y_2 \mid X = x_2) \ldots
\]

\[
= \prod_{i=1}^{n} P(Y = y_i \mid X = x_i)
\]

\[
= \prod_{i=1}^{n} \left\{ y_i f(\beta_0 + \beta_1 x_i) + (1 - y_i) (1 - f(\beta_0 + \beta_1 x_i)) \right\}
\]
Likelihood of data given model

\[ \ln \ a b = \ln a + \ln b \]
\[ \ln \ \prod_{i=1}^{n} a_i = \sum_{i=1}^{n} \ln a_i \]

Log likelihood:

\[ \ln \ P(Y=y \mid X=x) = \]
\[ \sum_{i=1}^{n} \ln \left\{ y_i f(\beta_0 + \beta_1 x_i) + (1-y_i) (1-f(\beta_0 + \beta_1 x_i)) \right\} \]
You can help fight COVID-19 by aiding research

https://covid.joinzoe.com/