Inf2 - Foundations of Data Science: Logistic regression

THE UNIVERSITY of EDINBURGH
informatics

FOUNDATIONS OF DATA SCIENCE

Announcements

- Week 4 workshop - we'll look at the paper that we'll be refer to in the exam

- Uses concepts from today's lecture!

- Solutions for Week 3 WS

A new unit:
The Maximum Likelihood Principle and Regression

Week 4: Logistic regression

Week 5: The maximum likelihood principle, and how
we can use it to derive linear, logistic and other types of
regression

Today

- Recap of Linear Regression

- Principle of Logistic Regression

- Interpretation of Logistic Regression coefficients

Wednesday:

- Multiple Logistic Regression

- Logistic Regression as a classifier

Inf2 - Foundations of Data Science: Recap of linear regression and classification

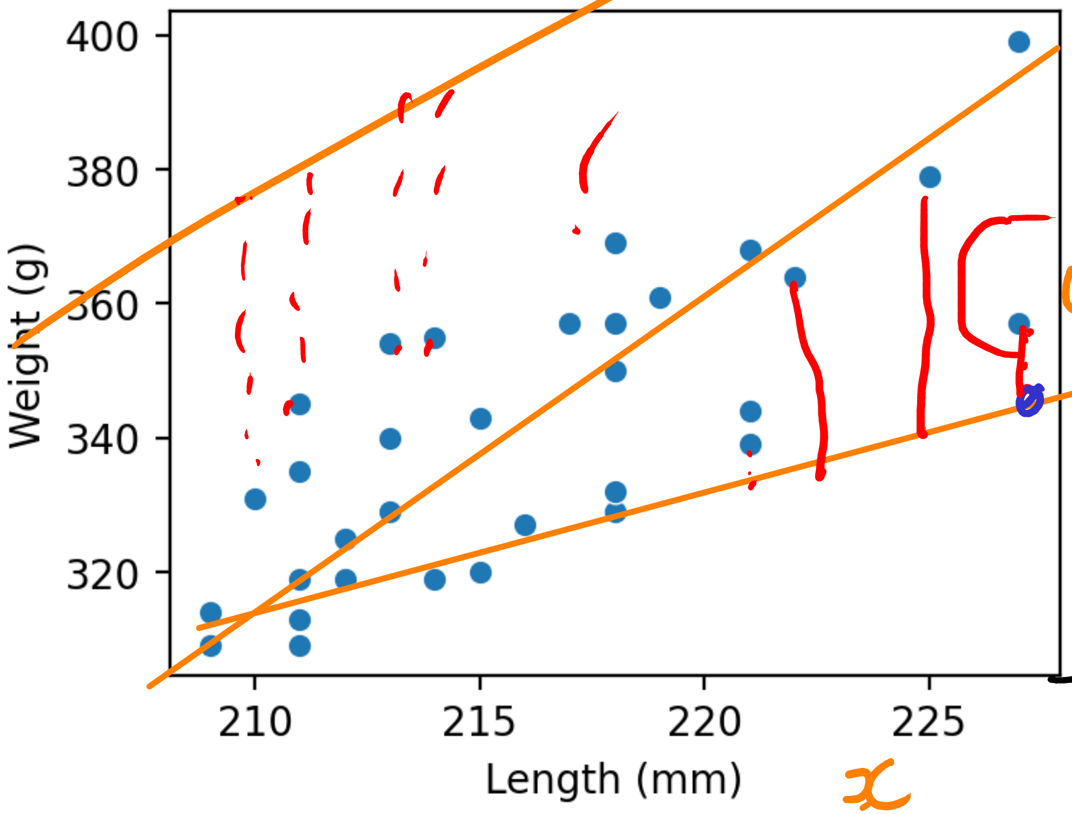
(Simple) Linear Regression

Minimise $\sum_{i=1}^{n} \varepsilon_i^2 = SSE$

- Given numeric predictor variable, predict a numeric response variable
- Regression coefficients set by minimising sum of squared errors between data and model (can be done analytically)

response

y



$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

$e_i = y_i - \hat{y}_i$

(x_i, y_i)

$\hat{y}_i$

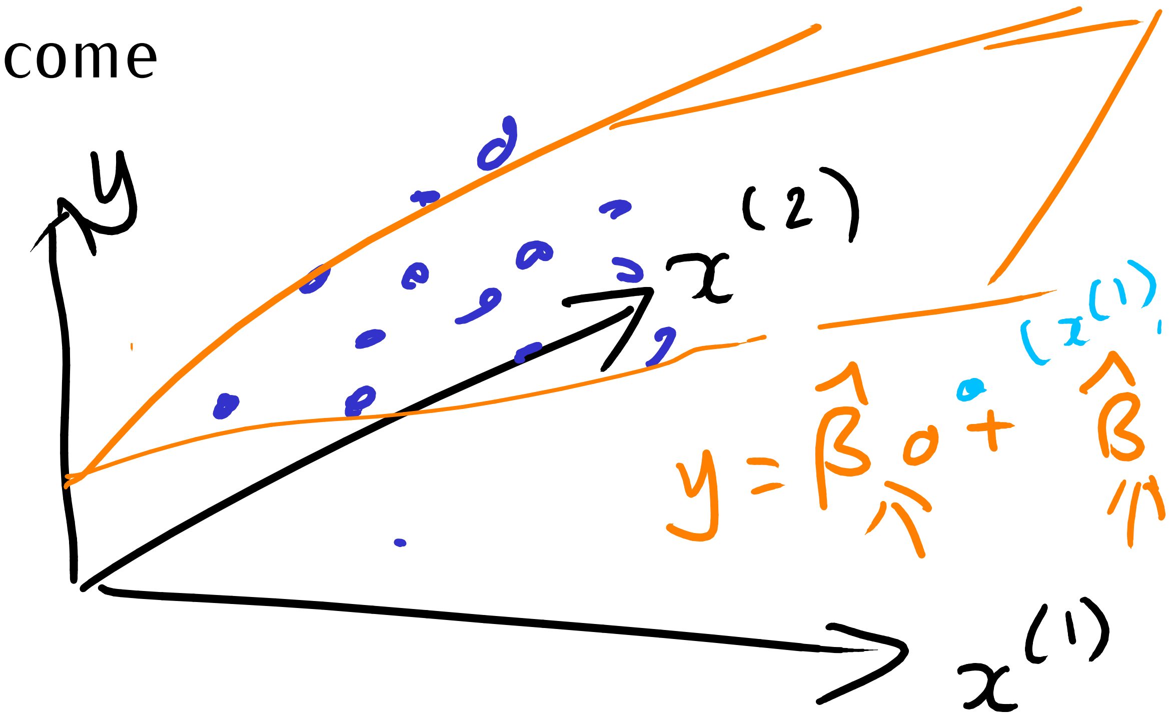
x

predictor

Multiple regression

- Multiple predictors
- Categorical predictors via one-hot encoding or indicator variables
- Can be used for prediction...
- ... or explanation, including"controlling for" variables not of interest

  => can use observational data to assess effect of treatment on outcome



Predict

$(x^{(1)}, x^{(2)}) \Rightarrow \hat{y}$

$y = \hat{\beta}_0 + \hat{\beta}_1 x^{(1)} + \hat{\beta}_2 x^{(2)}$

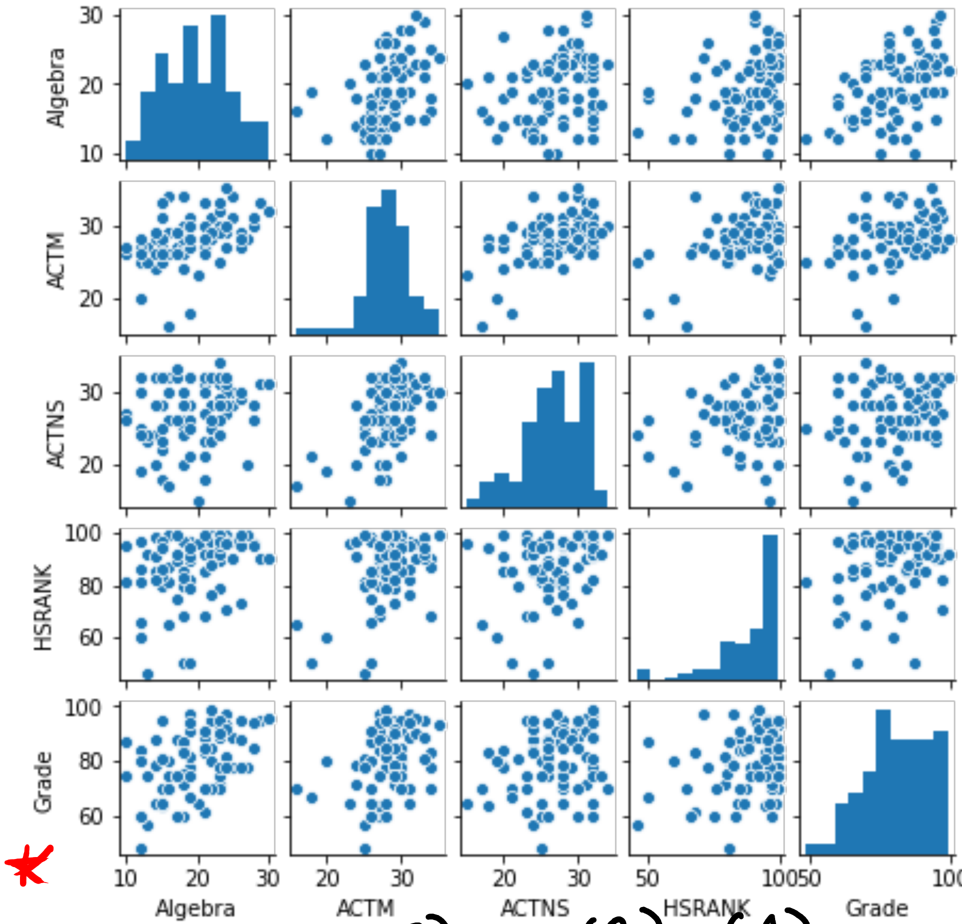
Confidence intervals and p-values of regression coefficients

Dep. Variable:	Grade	**R-squared:**	0.289
Model:	OLS	**Adj. R-squared:**	0.251
Method:	Least Squares	**F-statistic:**	7.622
Date:	Wed, 26 Oct 2022	**Prob (F-statistic):**	3.30e-05
Time:	09:42:47	**Log-Likelihood:**	-294.31
No. Observations:	80	**AIC:**	598.6
Df Residuals:	75	**BIC:**	610.5
Df Model:	4		
Covariance Type:	nonrobust		

	coef	**std err**	**t**	**P>|t|**	**[0.025**	**0.975]**
Intercept	36.1215	10.752	3.360	0.001	14.703	57.540
β_1 **Algebra**	0.9610	0.264	3.640	0.000	0.435	1.487
β_2 **ACTM**	0.2718	0.454	0.599	0.551	-0.632	1.175
ACTNS	0.2161	0.313	0.690	0.492	-0.408	0.840
HSRANK	0.1353	0.104	1.306	0.196	-0.071	0.342

$\hat{\beta}_1$ $\quad$ $\hat{\sigma}_{\hat{\beta}_1}$

***



$x^{(1)}$ $\quad x^{(2)}$ $\quad x^{(3)}$ $\quad x^{(4)}$ $\quad y$

$$T = \frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\hat{\beta}_1}}$$

Edge and Friedberg (1984)

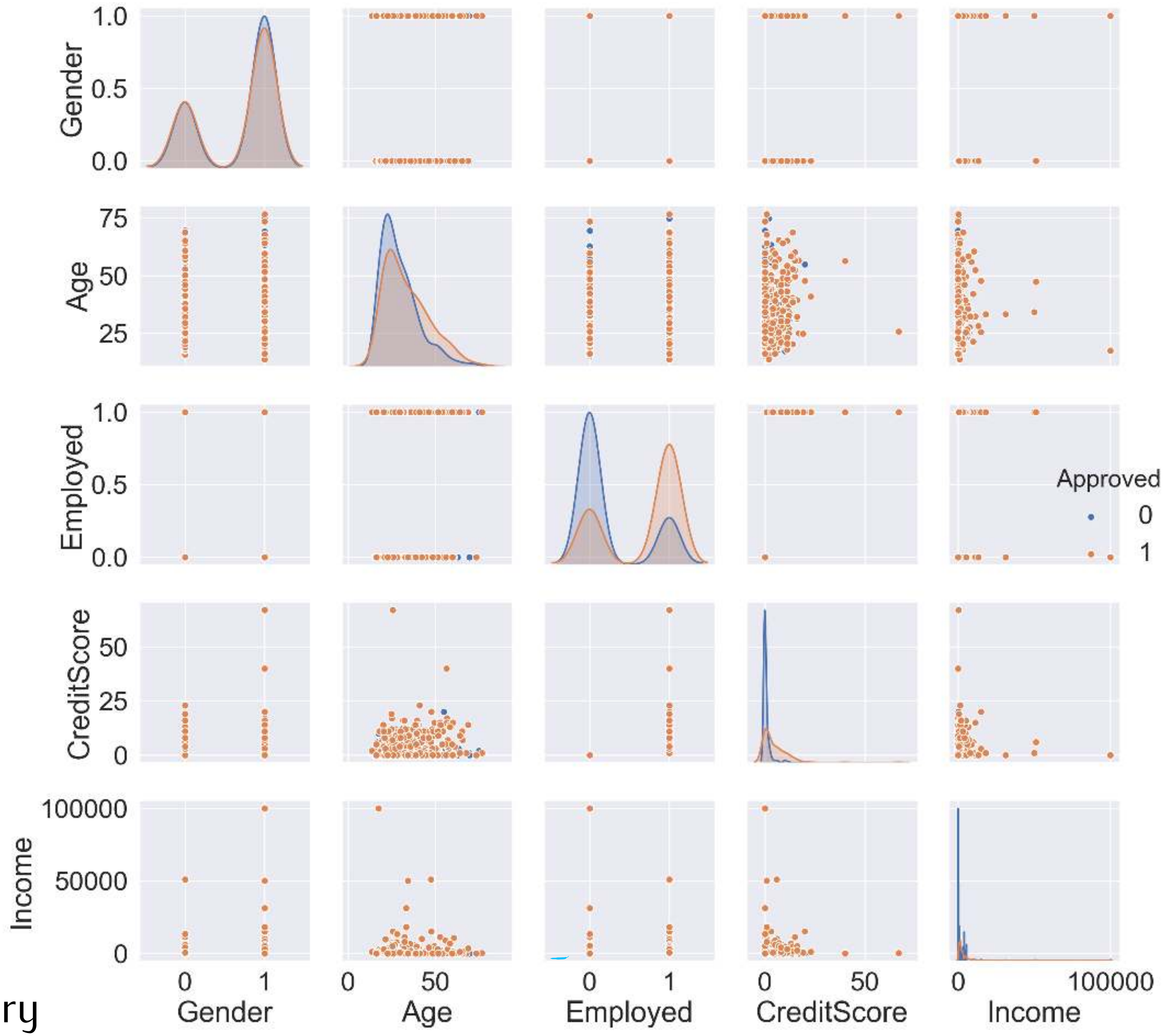
Inf2 - Foundations of Data Science: The principle of logistic regression

Supervised classification

Classifier predicts the label or class of an unseen point from features

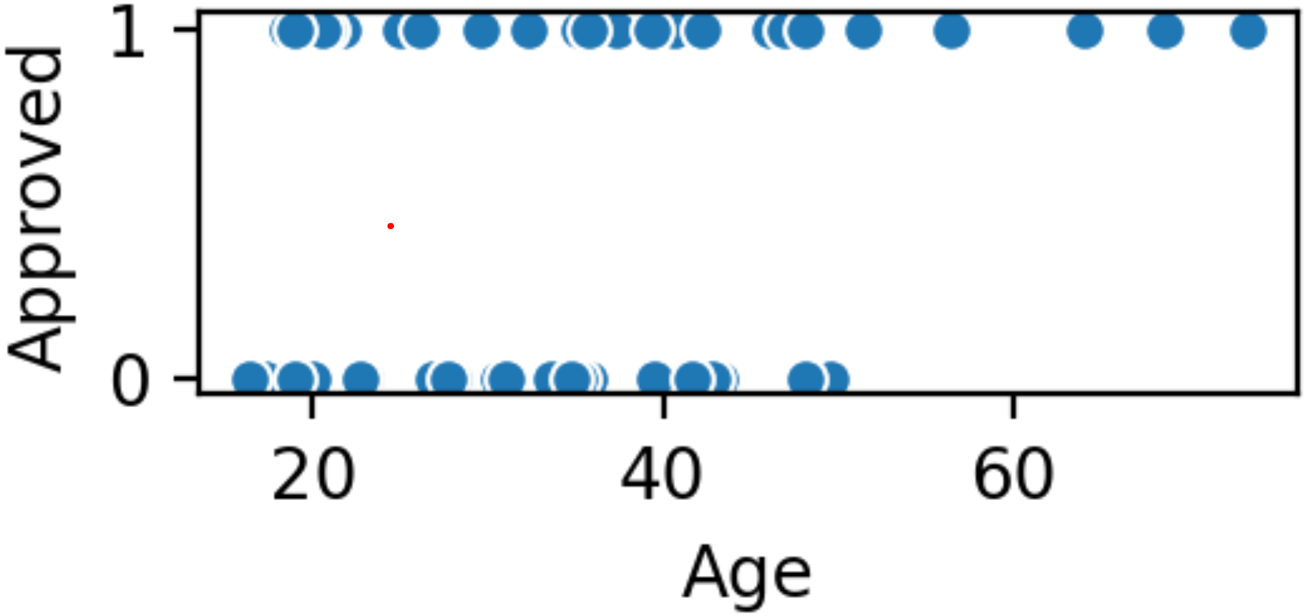
Binary (or dichotomous) response variable: Credit

Approved
Not approved

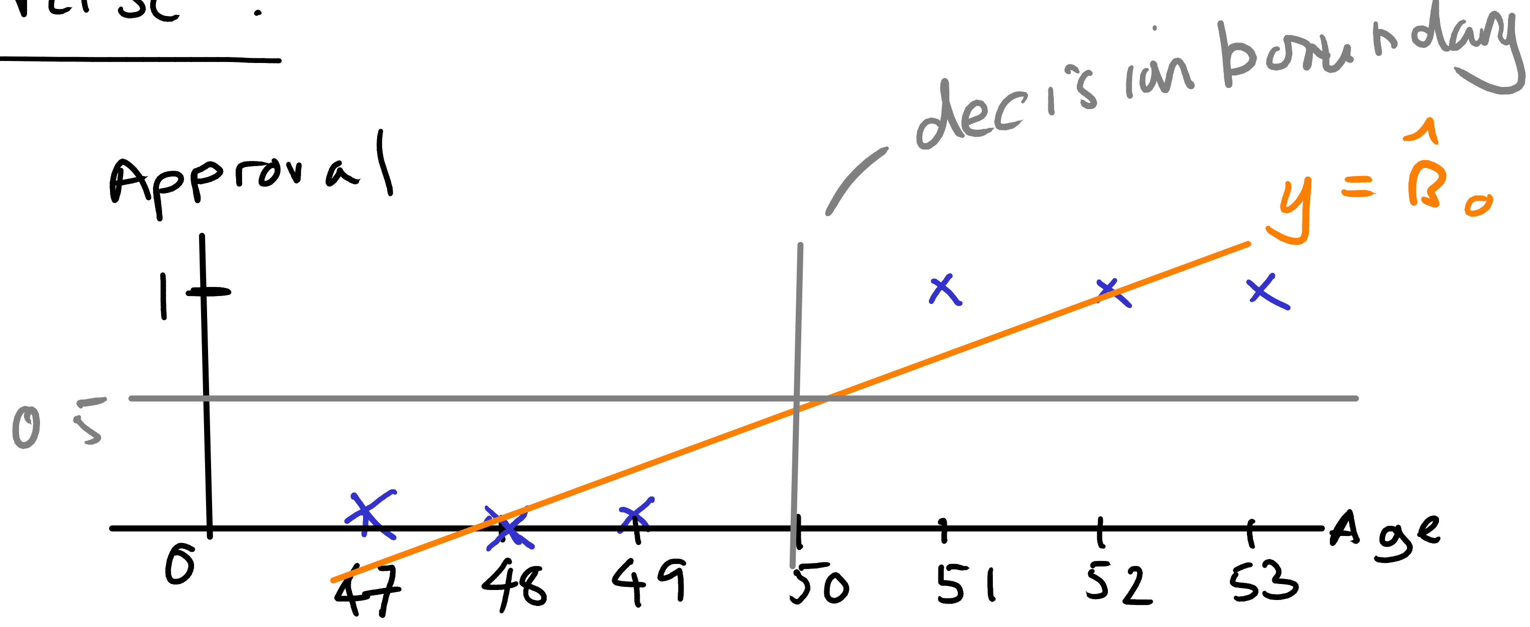


UCI Machine Learning Repository
https://archive.ics.uci.edu/dataset/27/credit+approval

Classification task on one continuous variable



Exercise :

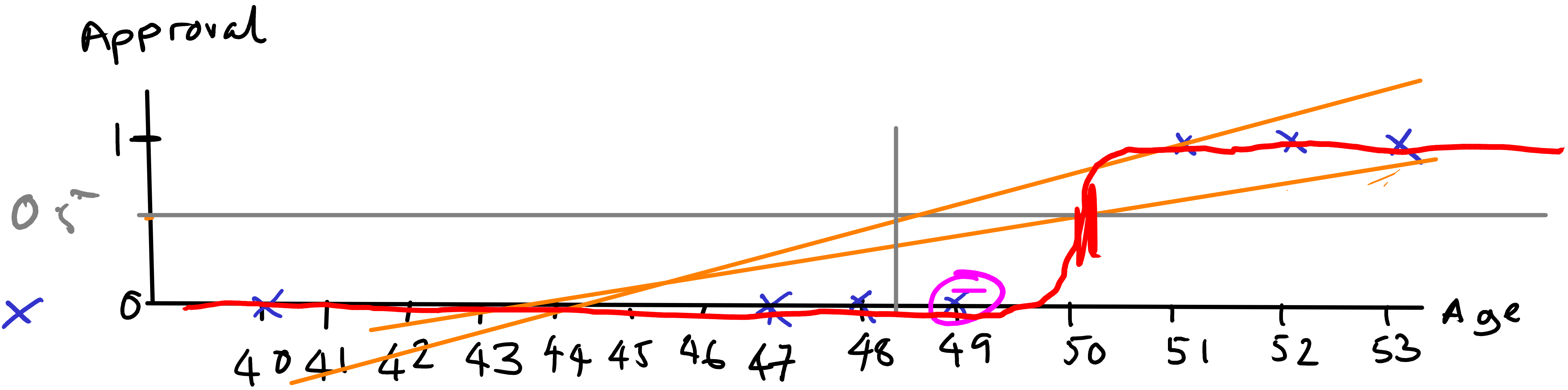


(a) Draw a Linear regression line through this toy data

(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 ?



(a) Draw Linear regression line through this toy data

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

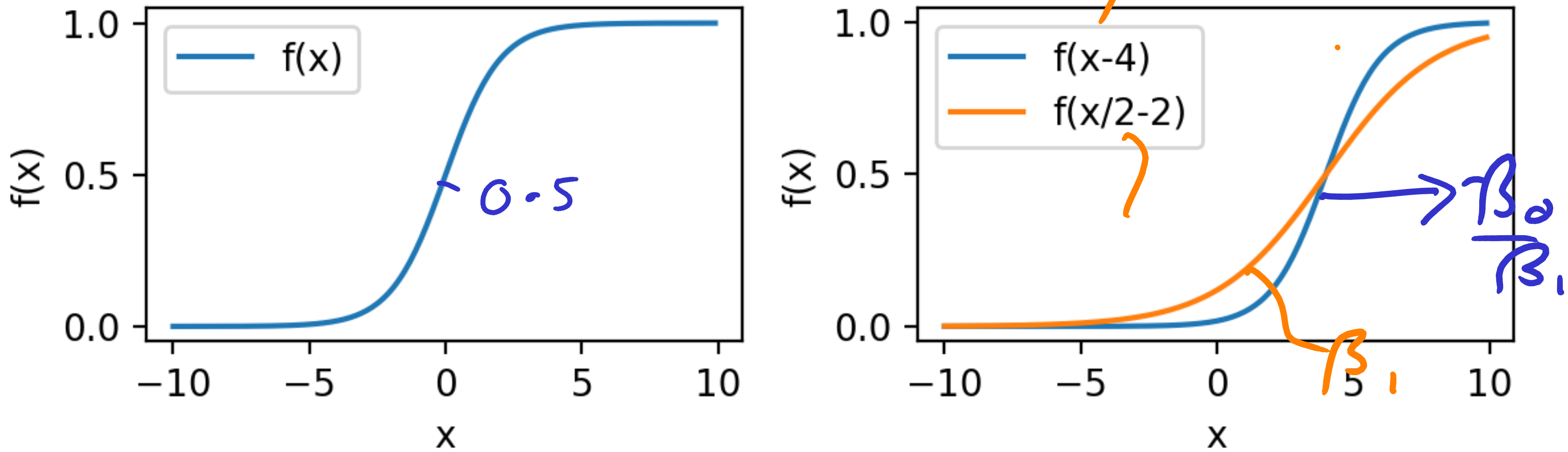
(c) Where is the decision boundary?

(d) How many classification errors are there?

Logistic function

$\beta_0 = 4$

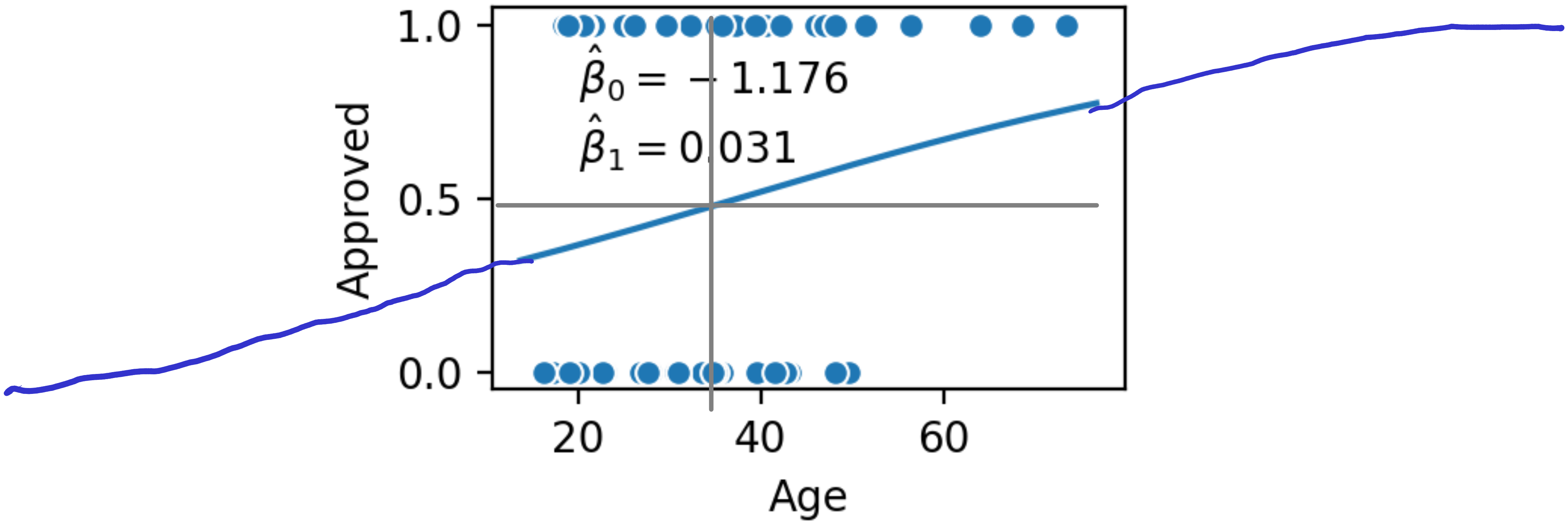
$\beta_1 = 1$



$$f(x) = \frac{e^{x}}{1+e^{x}} = \frac{1}{1+e^{-x}}$$

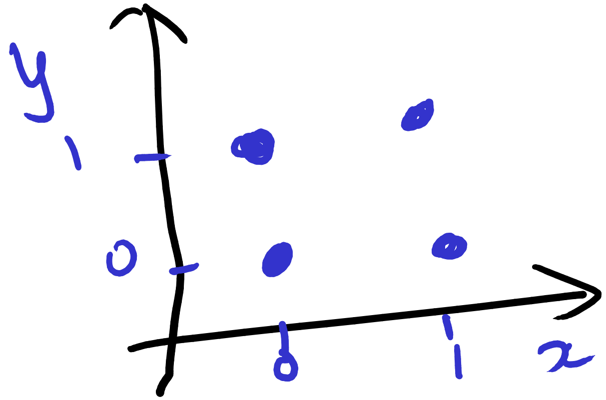
$$P(Y=1 \mid X=x) = f(\beta_0 + \beta_1 x) = \frac{1}{1 - e^{-\beta_0 - \beta_1 x}}$$

Application to continuous variable in credit example



Binary variables: odds and odds ratios

response Y $\quad P(Y=y \mid X=x)$



Employed	Approved	Not approved	Approval odds
0	p 0.25	0.75	0.34
1	0.71	0.29	2.42

OR

X predictor

$\text{Odds}(\text{App.} \mid \text{Emp.}) = \dfrac{0.25}{0.75} = 0.34$

$y \in \{\text{"Not approved"}, \text{"Approved"}\}$

$x \in \{\text{"Not Emp."}, \text{"Emp."}\}$

$$\text{Odds}(\text{Success}) = \frac{P(\text{Sucess})}{P(\text{Failure})} = \frac{P(\text{Sucess})}{1 - P(\text{Success})}$$

Odds Ratio $\quad OR(x) = \dfrac{\text{Odds}(\text{success} \mid x = \text{True})}{\text{Odd}(\text{sucess} \mid x = \text{False})}$

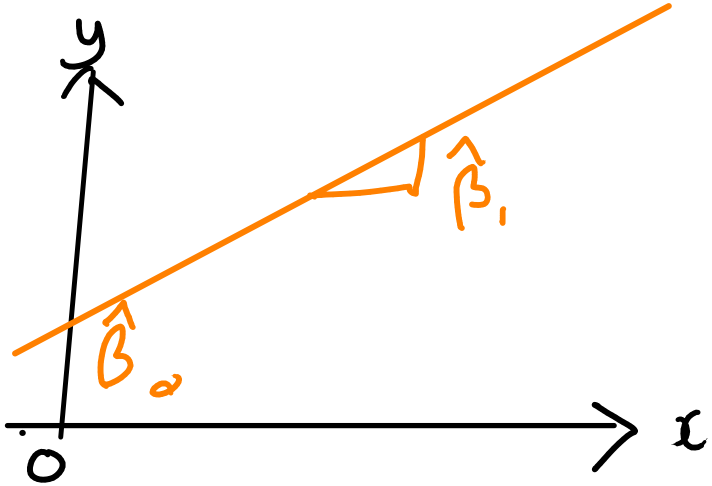
$OR(\text{Employed}) = \dfrac{2.42}{0.34} = 7.02$

Effect size $= 602\%$

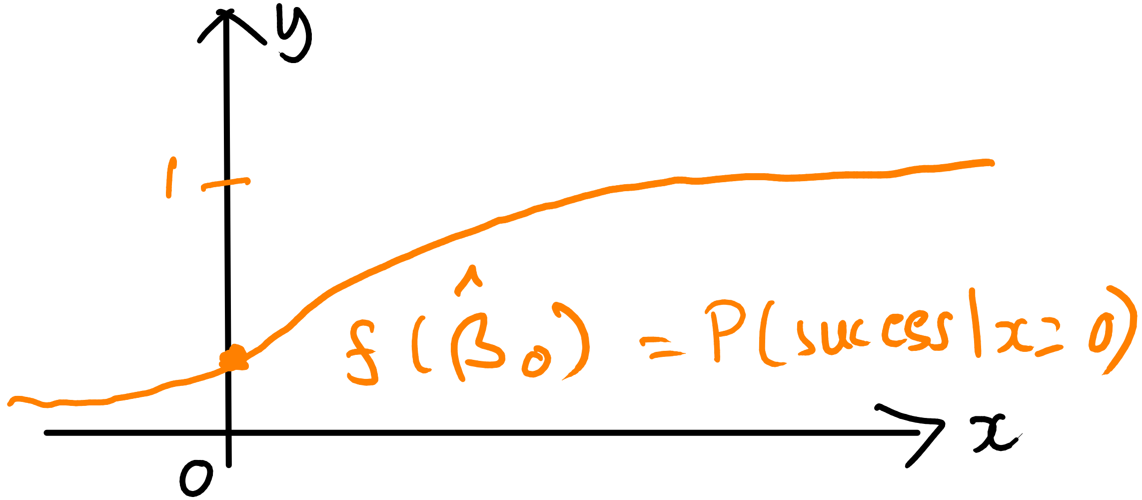
Inf2 - Foundations of Data Science: Interpretation of logistic regression coefficients

Interpretation of $\hat{\beta}_0$

Lin reg



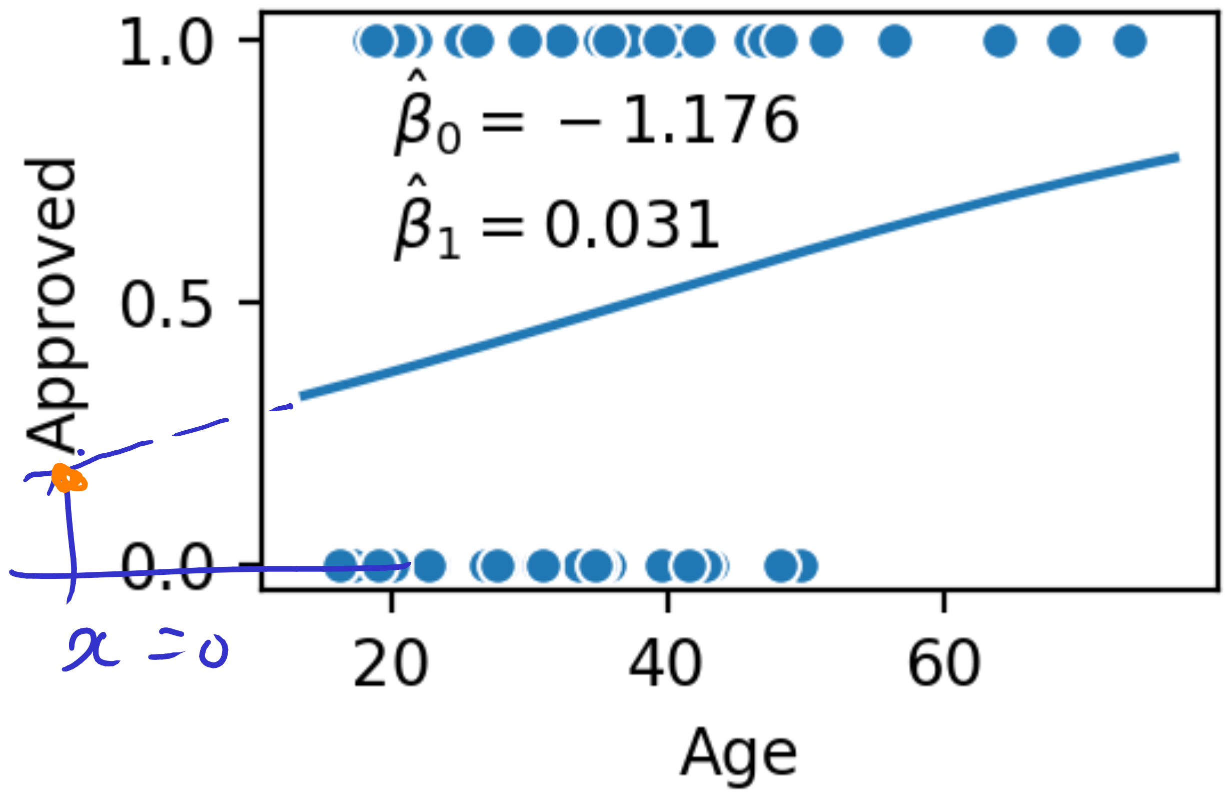
Log. reg



$$f(\hat{\beta}_0 + \hat{\beta}_1 \cdot 0) = f(\hat{\beta}_0) = \frac{1}{1+e^{-\hat{\beta}_0}}$$

$f(\hat{\beta}_0) = f(-1.176)$

$= 0.236$

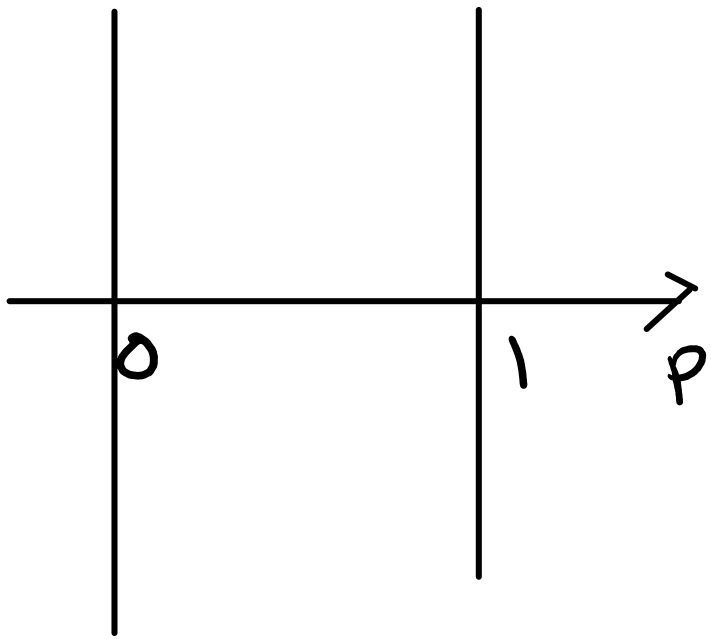


Log odds

$$\text{Log Odds(Success)} = \underset{\log_e}{\ln} \frac{P(\text{Success})}{P(\text{Failure})}$$

$$= \ln \frac{P(\text{Success})}{1 - P(\text{Success})} = \ln \frac{p}{1-p}$$

p	odds	Log odds
0.5	1	0
>0.5	>1	>0
<0.5	<1	<0
1	$\to \infty$	$\to \infty$
0	0	$\to -\infty$



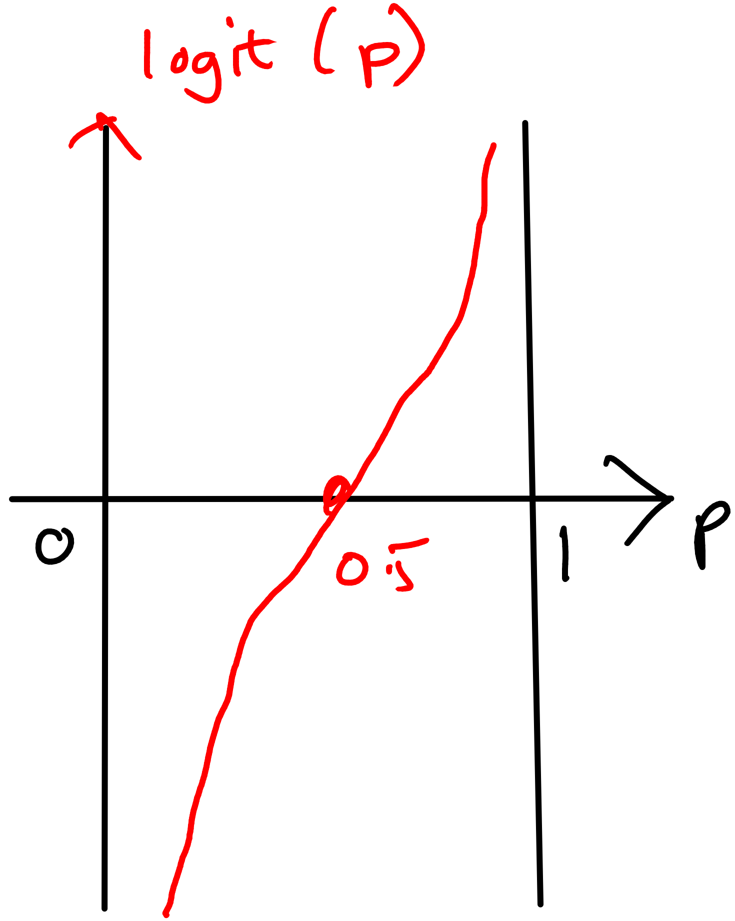
Logit scale

Log odds $+1$ $\Rightarrow$ odds increase by factor e

logistic unit = logit

$\hat{\beta}_0 = -1.176$ logits

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



Logistic regression in terms of log odds

Success $\quad P(Y=1 \mid x) = f(\beta_0 + \beta_1 x) = \dfrac{1}{1 + e^{-\beta_0 - \beta_1 x}}$ ①

Failure $\quad P(Y=0 \mid x) = 1 - f(\beta_0 + \beta_1 x) = 1 - \dfrac{1}{1 + e^{-\beta_0 - \beta_1 x}}$

$$= \frac{e^{-\beta_0 - \beta_1 x}}{1 + e^{-\beta_0 - \beta_1 x}} \quad ②$$

① / ②

Odds $\quad \dfrac{P(Y=1 \mid x)}{P(Y=0 \mid x)} = \dfrac{1}{e^{-\beta_0 - \beta_1 x}} = \underline{\underline{e^{\beta_0 + \beta_1 x}}}$

Log odds $\quad \ln \dfrac{P(Y=1 \mid x)}{P(Y=0 \mid x)} = \beta_0 + \beta_1 x = \operatorname{logit}(P(Y=1 \mid x))$

Interpretation of $\hat{\beta}_1$

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

$$= e^{\hat{\beta}_0} e^{\hat{\beta}_1 x}$$

$x = \{0, 1\}$

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

$$= e^{\hat{\beta}_1 x}$$

$$\log OR(x) = \hat{\beta}_1$$

$\hat{\beta}_1 = 0.03$

Credit e.g.

$$OR(\text{Age}) = e^{0.03} \simeq 1.03$$

Preview

- Almost all pieces now in place to understand most of target paper

- Indicator variables

- Odds ratios

- Confidence intervals

- Last piece: Multiple logistic regression (next time)

$\beta_0 + \beta_1 x^{(1)} + \beta_2 x^{(2)} = \text{logit}(P(Y=1|x))$

Summary

- Logistic regression as a classification task

- Transforming linear regresssion into logistic regression: The sigmod function

- Odds, log odds, and odds ratios