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

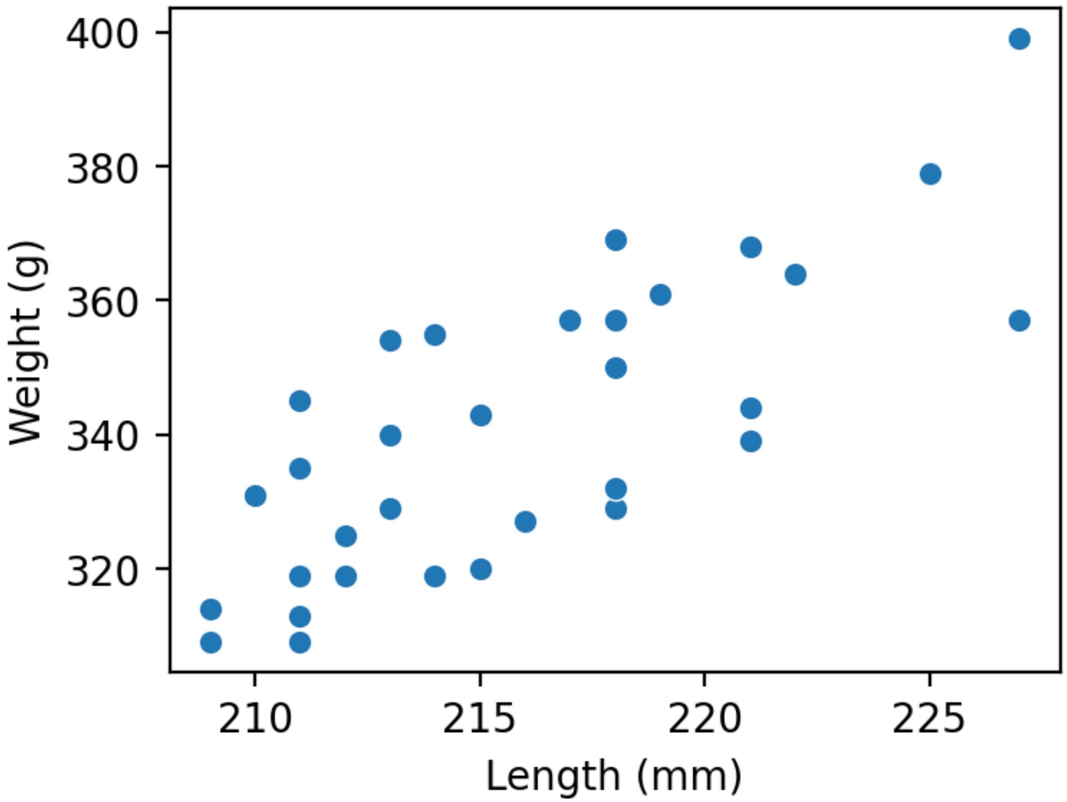
Inf2 - Foundations of Data Science: Recap of linear regression and classification

THE UNIVERSITY *of* EDINBURGH
informatics

FOUNDATIONS OF DATA SCIENCE

(Simple) Linear Regression

- 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)



Multiple regression

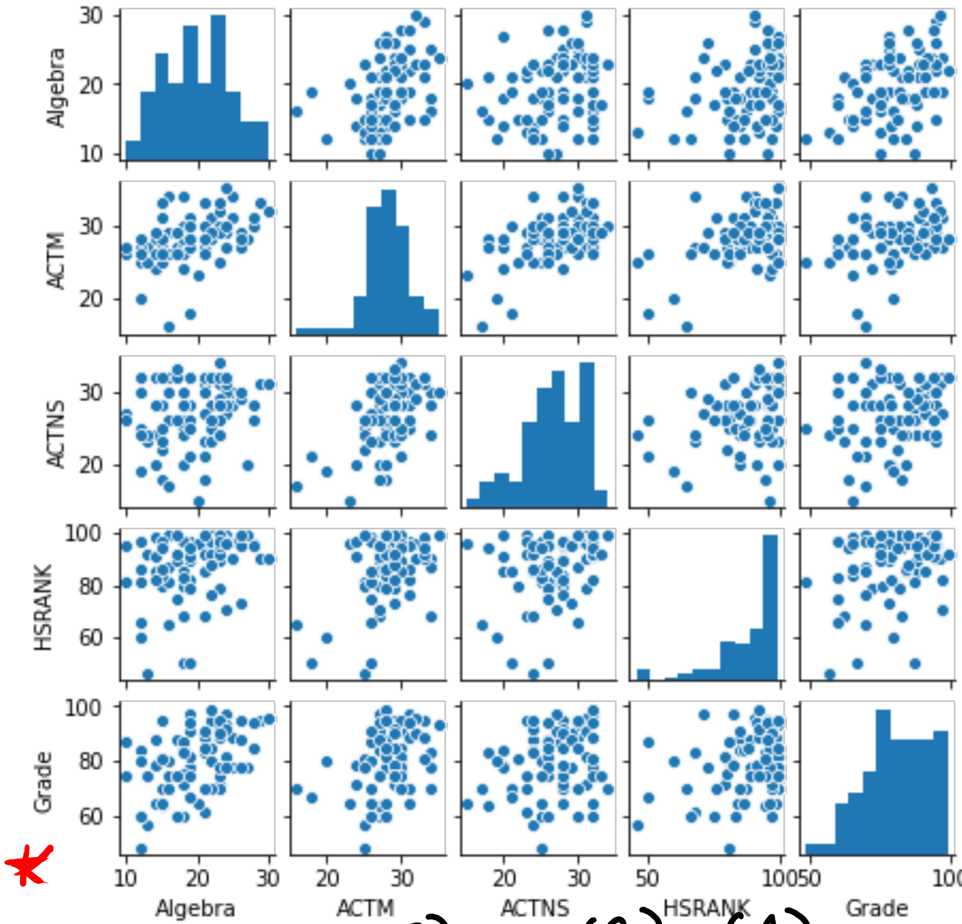
- Multiple predictors
- Categorical predictors via one-hot encoding or indicator variables
- Can be used for prediction...
- ... or explanation, including"controlling for" variables not of interest
  => can use observational data to assess effect of treatment
  on outcome

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
Algebra	0.9610	0.264	3.640	0.000	0.435	1.487
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

β_1 β_2 β_1 $\hat{\sigma}_{\hat{\beta}_1}$ ***



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

Edge and Friedberg (1984)

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

THE UNIVERSITY of EDINBURGH
informatics

FOUNDATIONS OF DATA SCIENCE

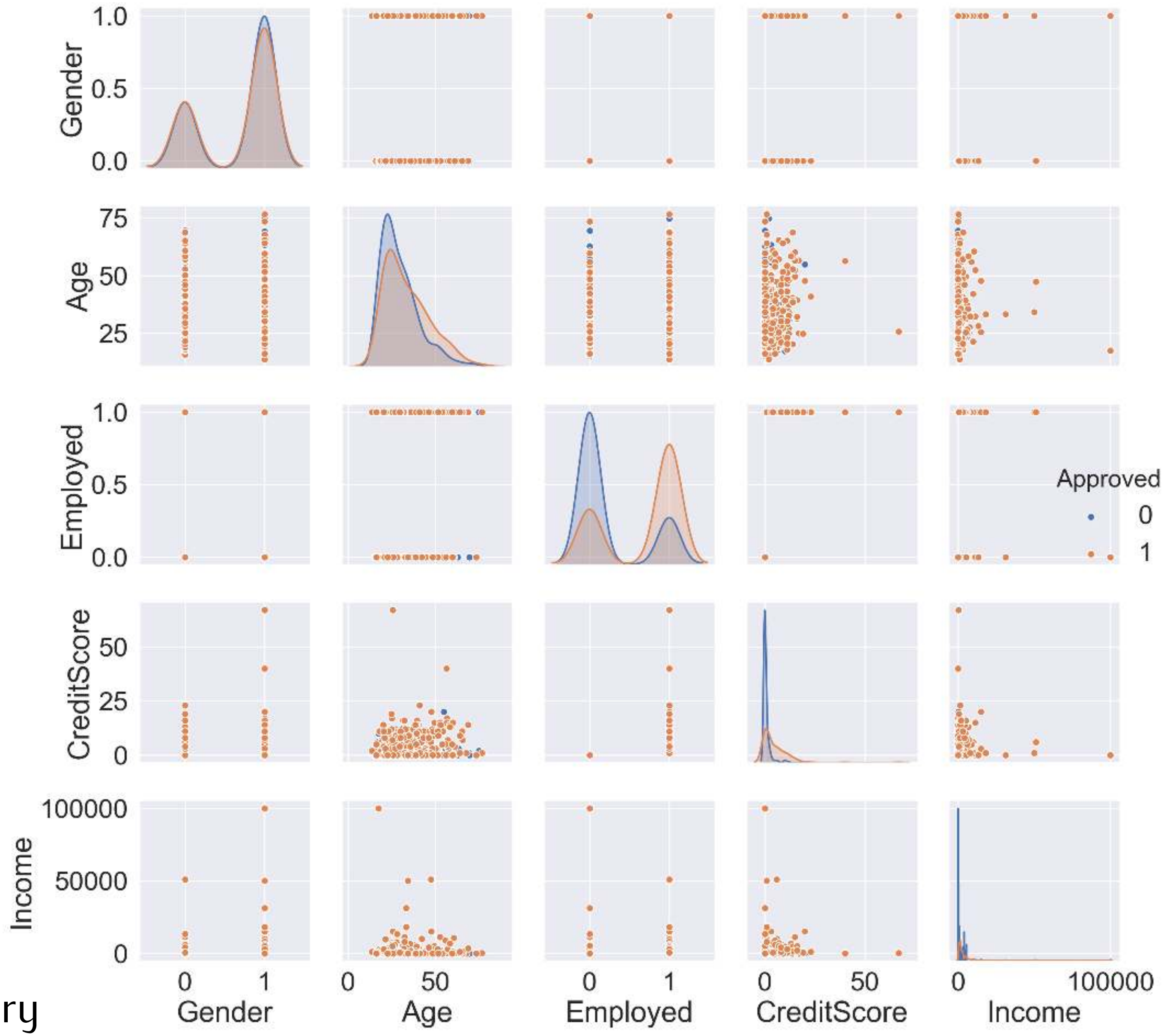
Supervised classification

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

Binary (or dichotomous) response variable: Credit

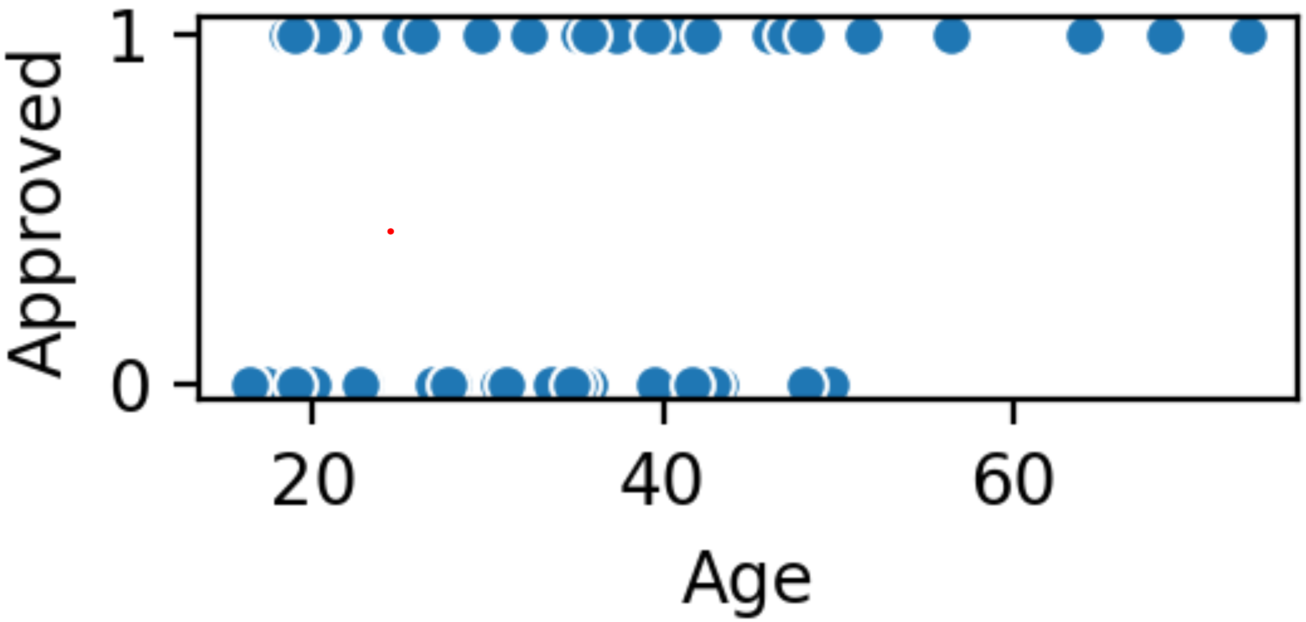
Approved

Not approved

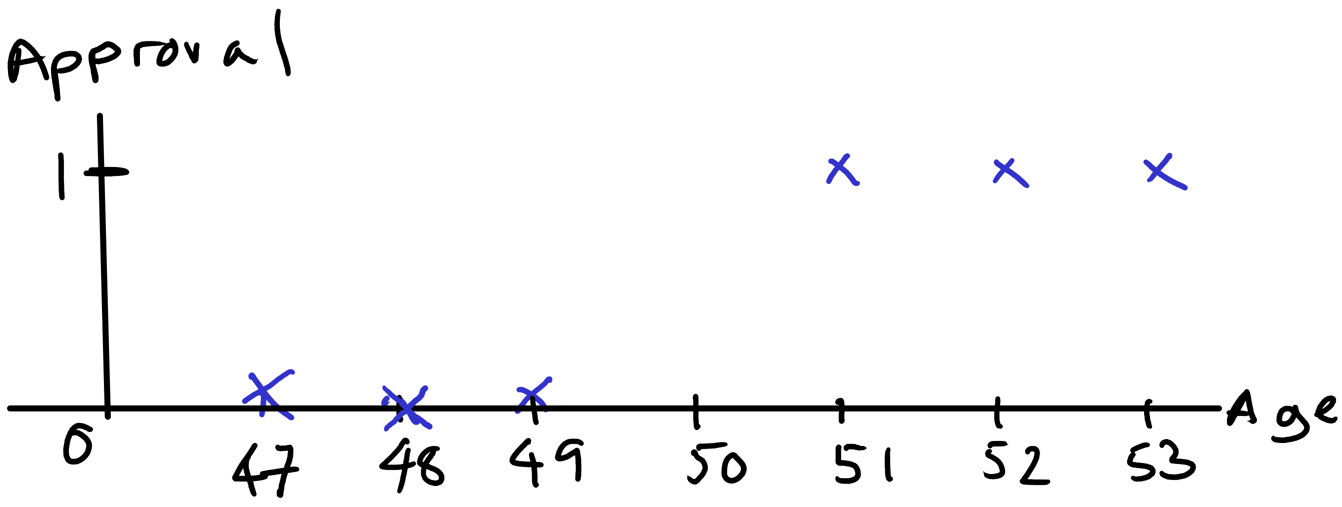


UCI Machine Learning Repository
https://archive.ics.uci.edu/dataset/27/credit+approval

Classification task on one continuous variable



Exercise :

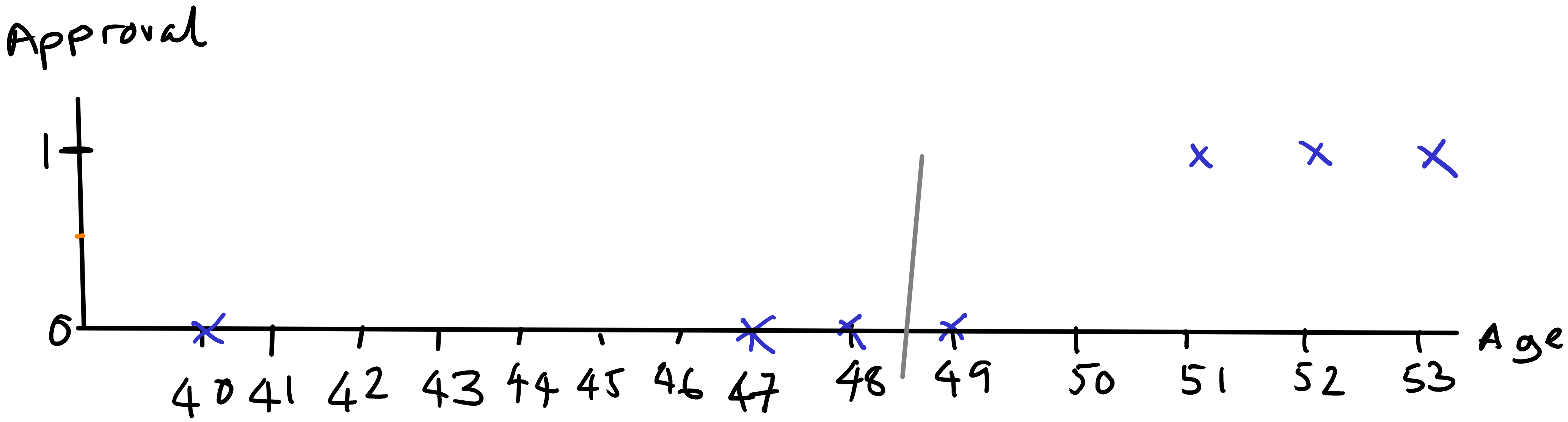


(a) Draw a Linear regression line through this toy data

(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 ?



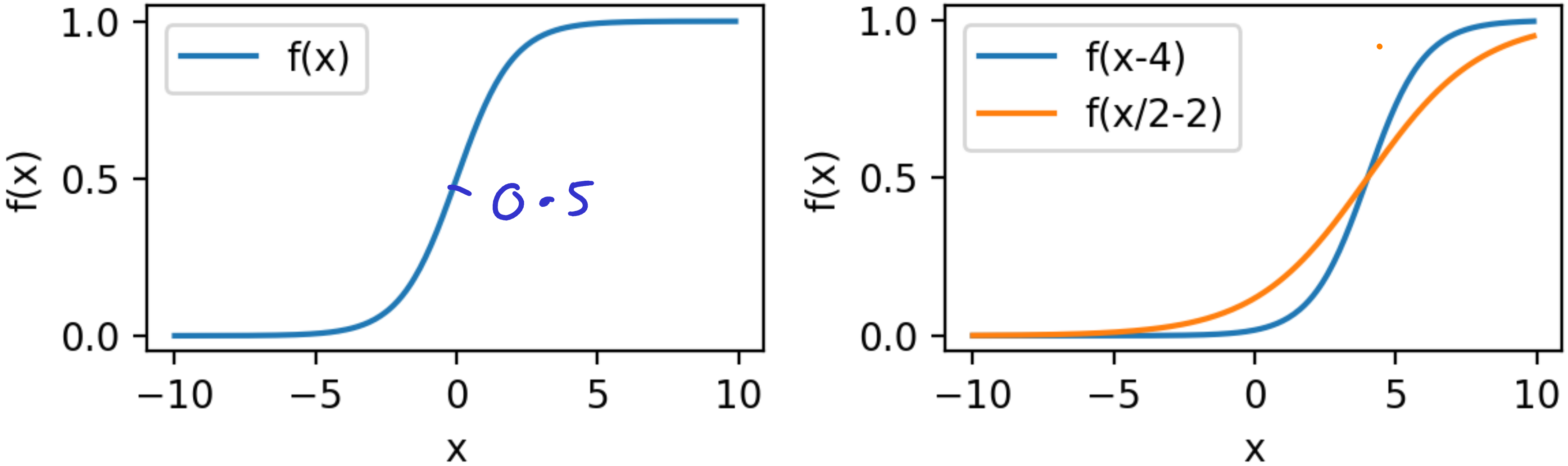
(a) Draw Linear regression line through this toy data

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

(c) Where is the decision boundary?

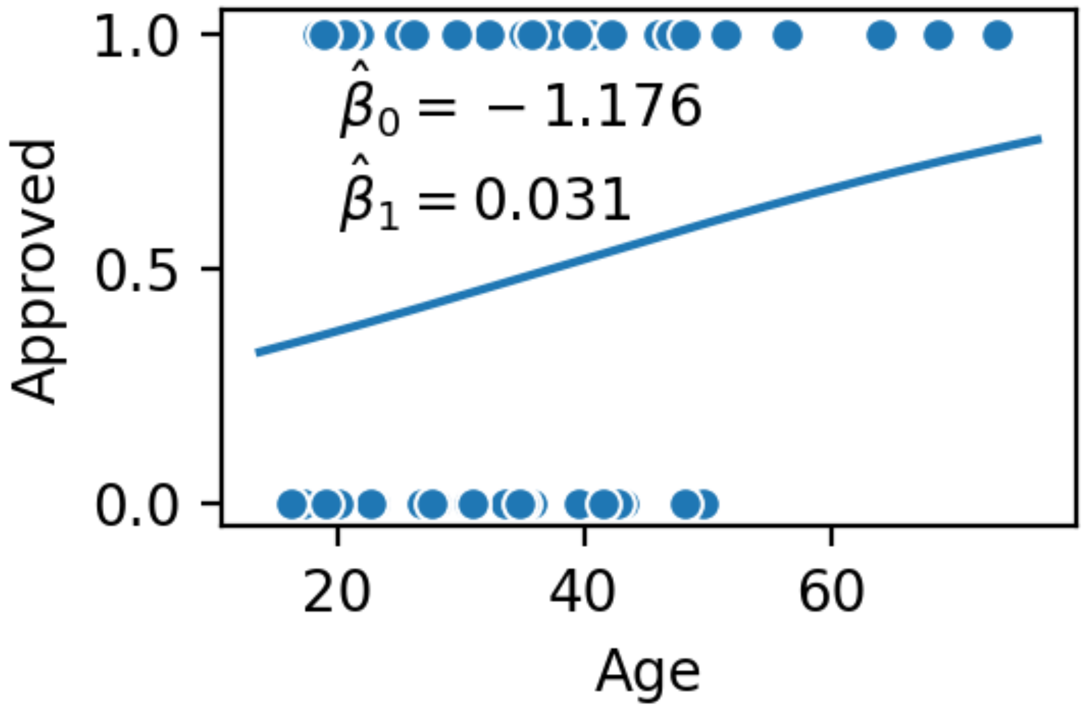
(d) How many classification errors are there?

Logistic function



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

Application to continuous variable in credit example



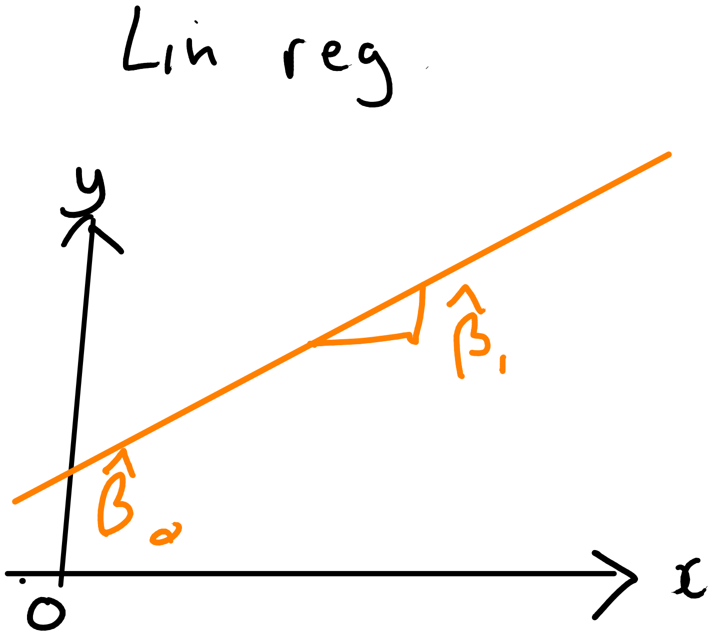
Binary variables: odds and odds ratios

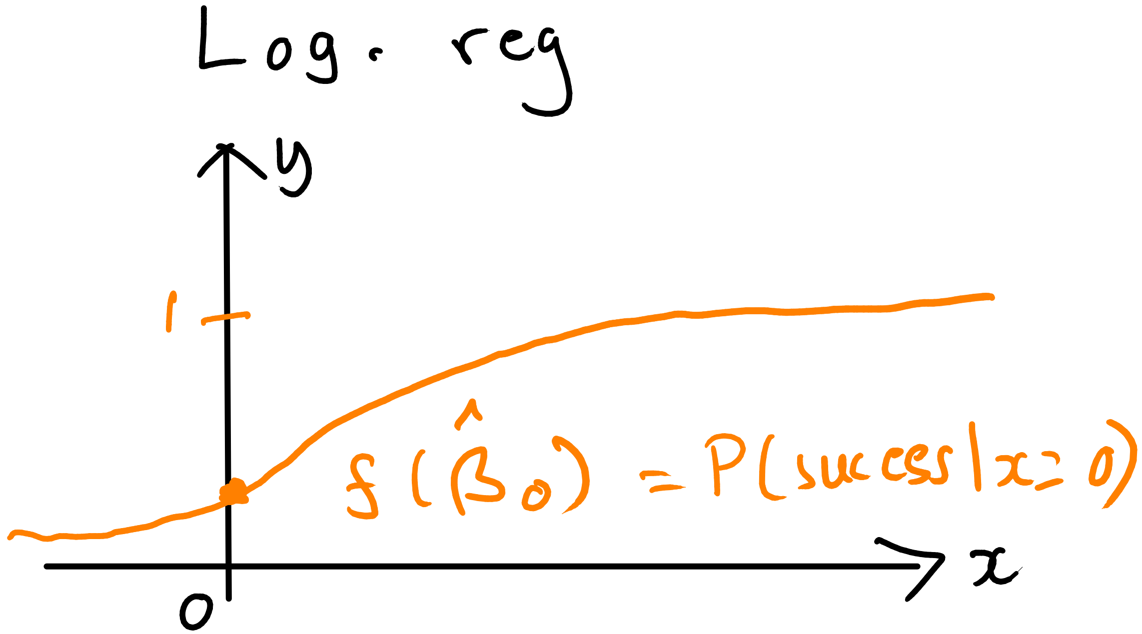
Employed	Approved	Not approved	Approval odds
0	0.25	0.75	0.34
1	0.71	0.29	2.42

$y \in \{$ "Not approved", "Approved" $\}$
$x \in \{$ "Not Emp.", "Emp." $\}$

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

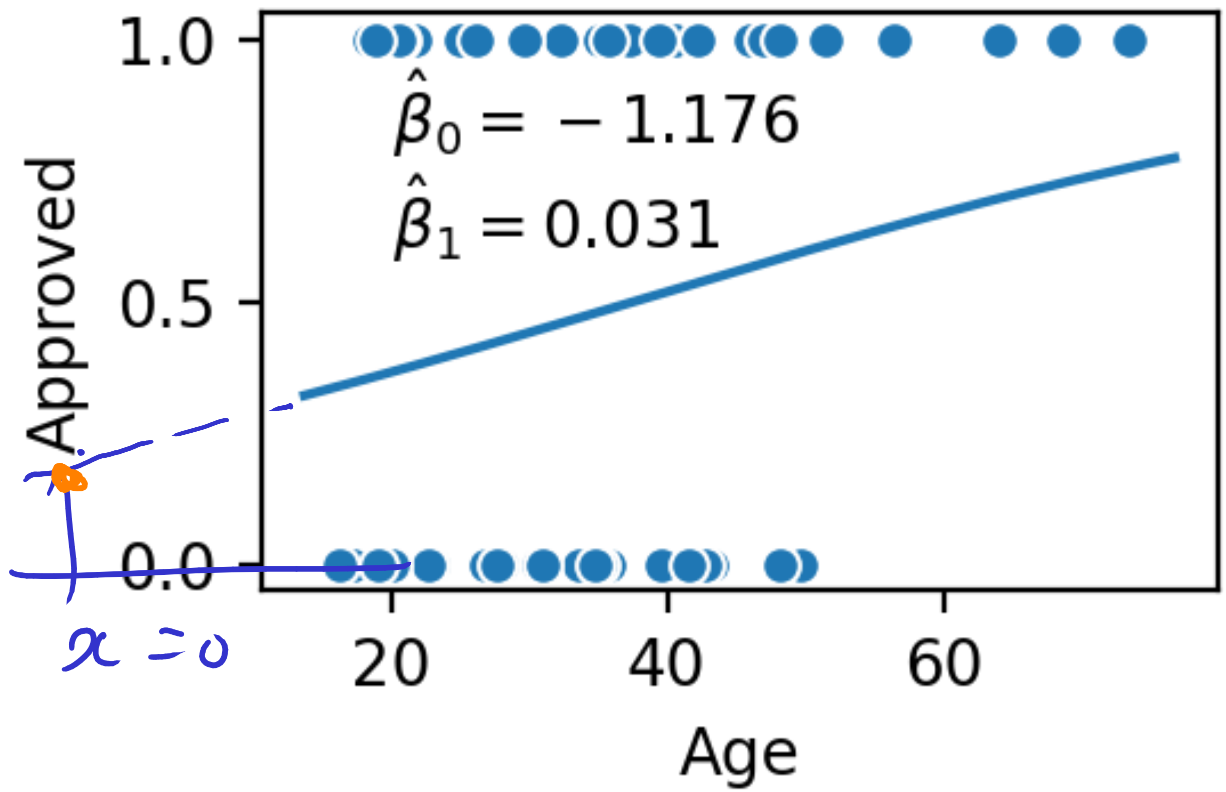
Interpretation of $\hat{\beta}_0$





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

$= 0.236$



Log odds

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

($\ln$ ← $\log_e$)

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

p	odds	Log odds
0.5		
>0.5		
<0.5		
1		
0		

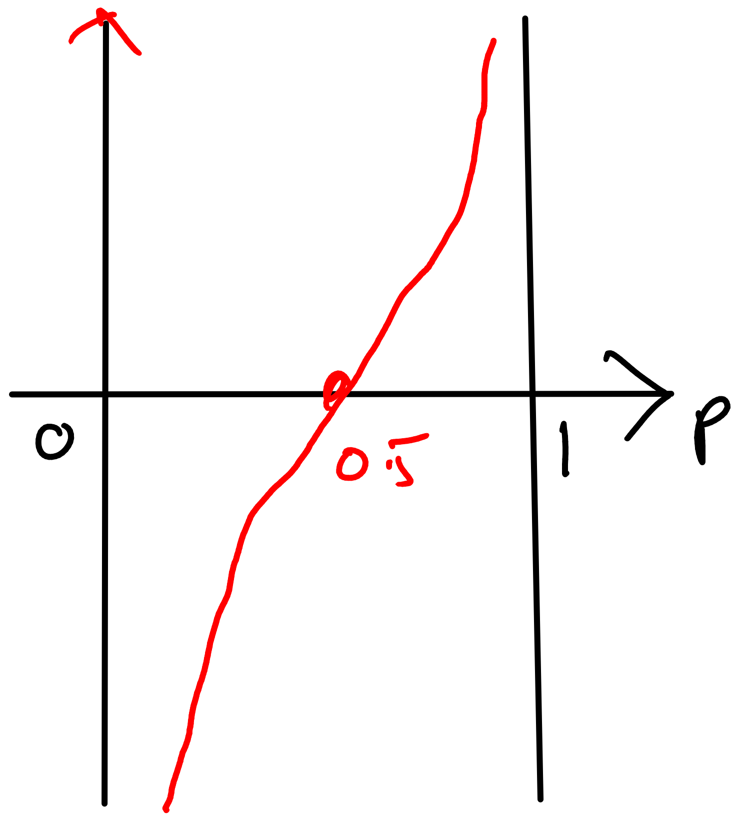
Logit scale

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

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

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

① / ②

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

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

Interpretation of $\hat{\beta}_1$

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

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

$x = \{0, 1\}$

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)

Summary

- Logistic regression as a classification task

- Transforming linear regresssion into
  logistic regression: The sigmod function

- Odds, log odds, and odds ratios