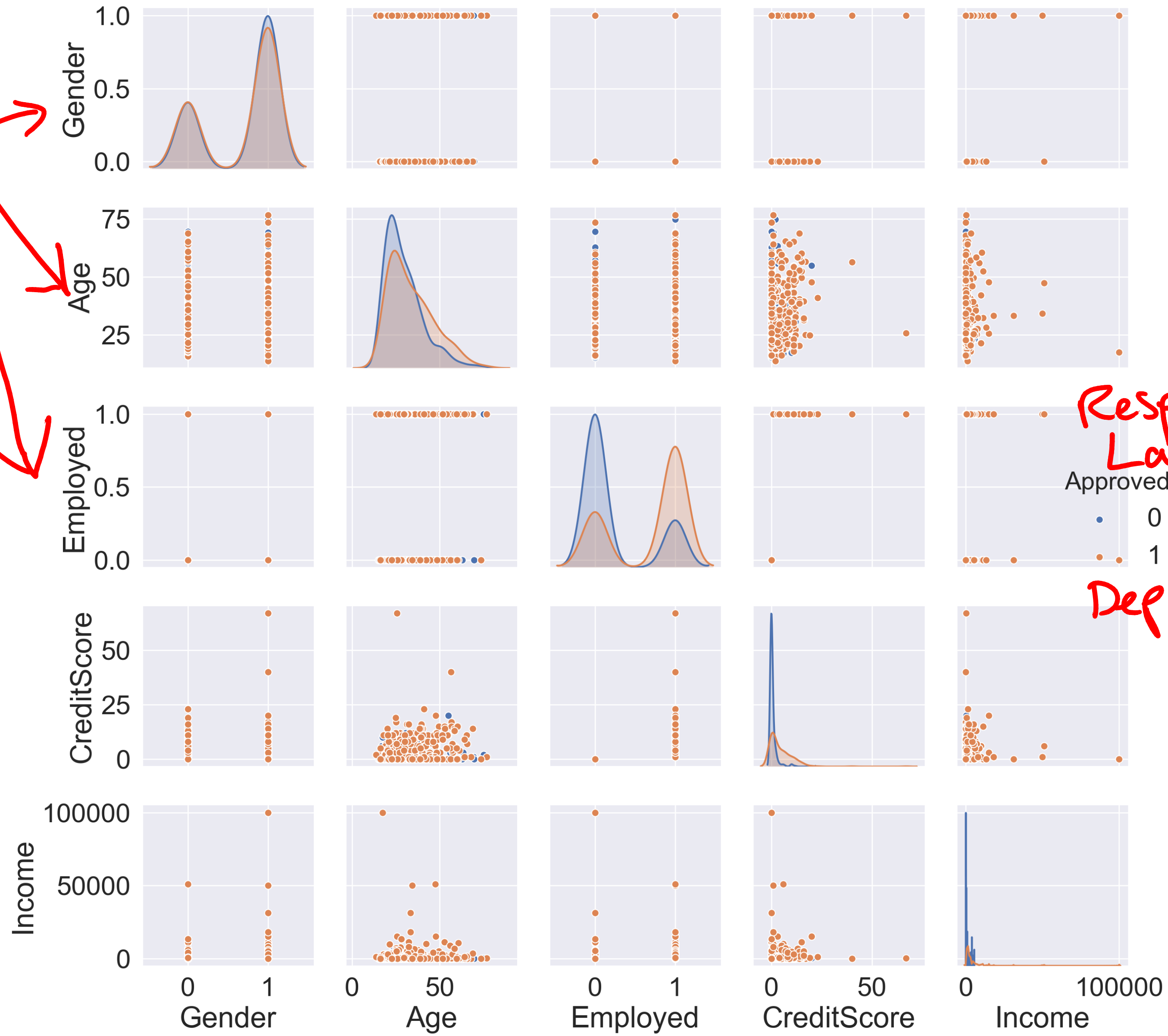
Foundations of Data Science: Logistic regression - Principle of logistic regression

Supervised classification

Indep. vars
Features
predictor vars

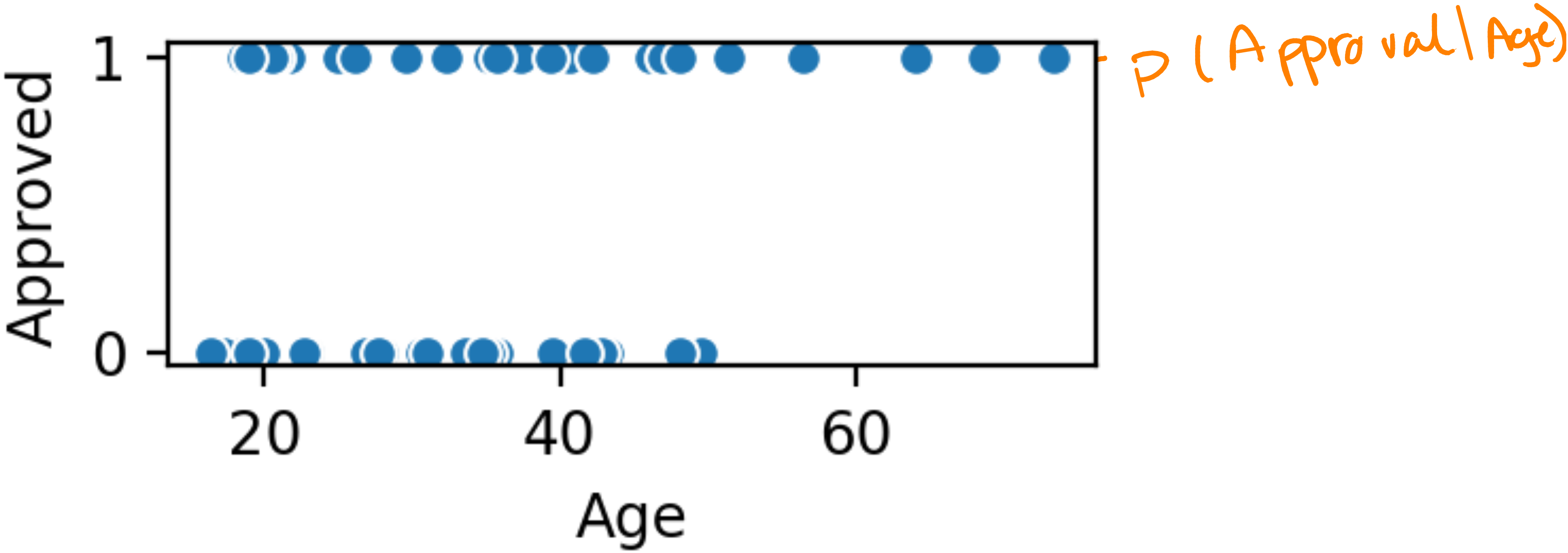


Resp.
Label

Dep var.

Logistic regression
predict
probability

Logistic regression task on one continuous variable



Binary variables: odds and odds ratios

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

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

odds ratio

$OR(\text{Employed}) = \frac{2.41}{0.34} = 7.09$

Effect size 609%

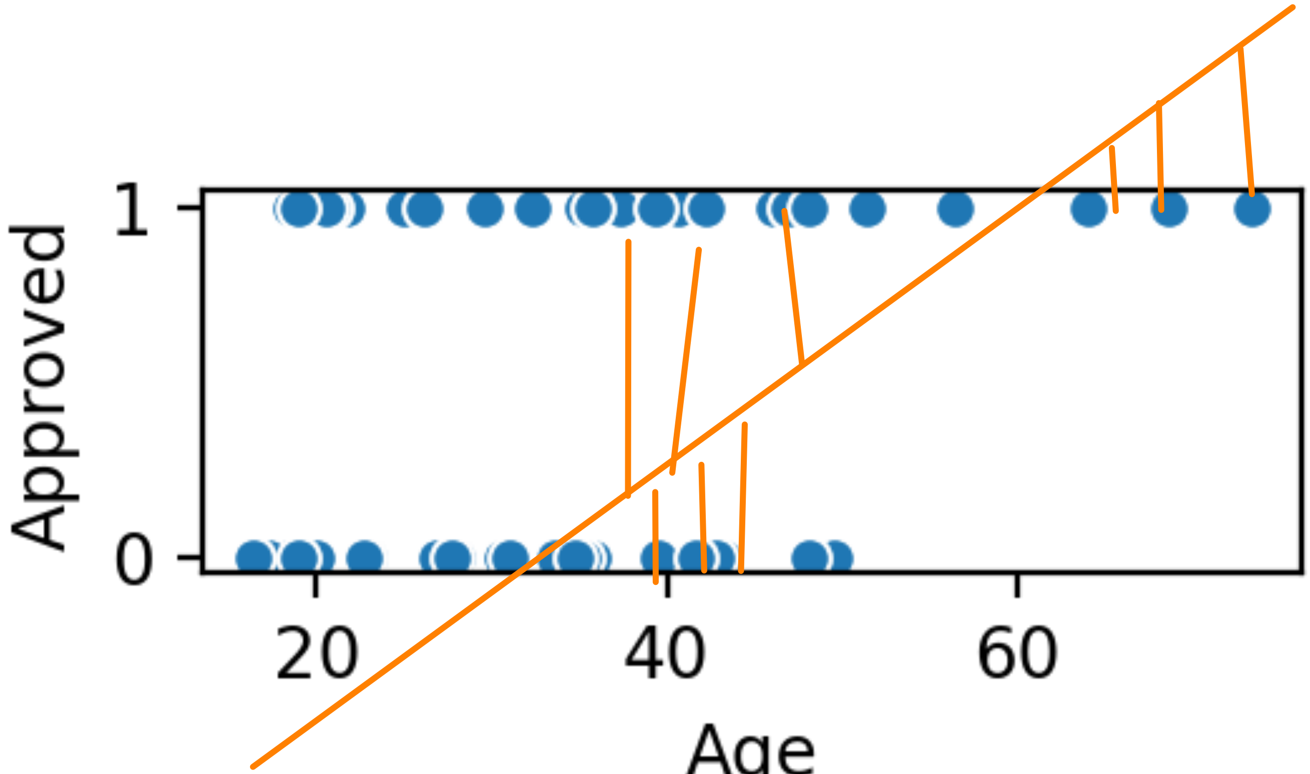
$y \in \{\text{Approved}, \text{Not approved}\}$

$x \in \{\text{Employed}, \text{Not Employed}\}$

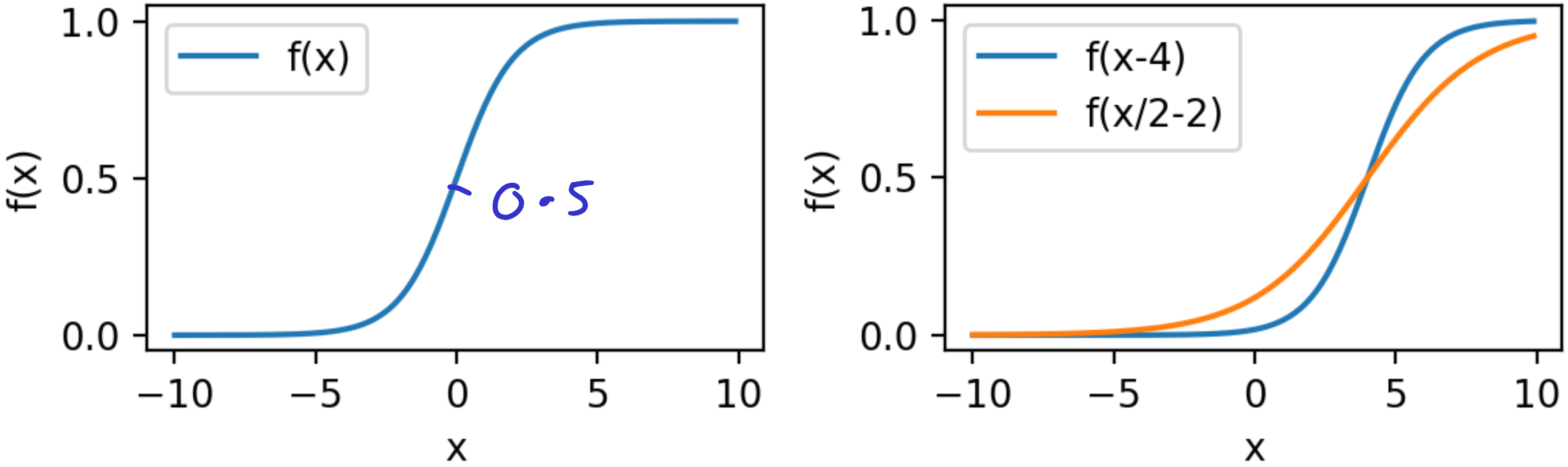
$$\text{Odds}(\text{Success}) = \frac{P(\text{Success})}{P(\text{Failure})} = \frac{P(\text{Success})}{1 - P(\text{Success})}$$

Odds ratio $OR(x) = \dfrac{\text{Odds}(\text{Success} \mid x = \text{True})}{\text{Odd}(\text{Success} \mid x = \text{False})}$

Principle of logistic regression with one variable



Logistic function



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

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

$\hat{\beta}_0 \quad \hat{\beta}_1$

coeffs

Application to continuous variable in credit example