

Regression

Classification recap



Pyronia tithonus
Maniola jurtina
Pieris brassicae



Spam
Not spam

Regression

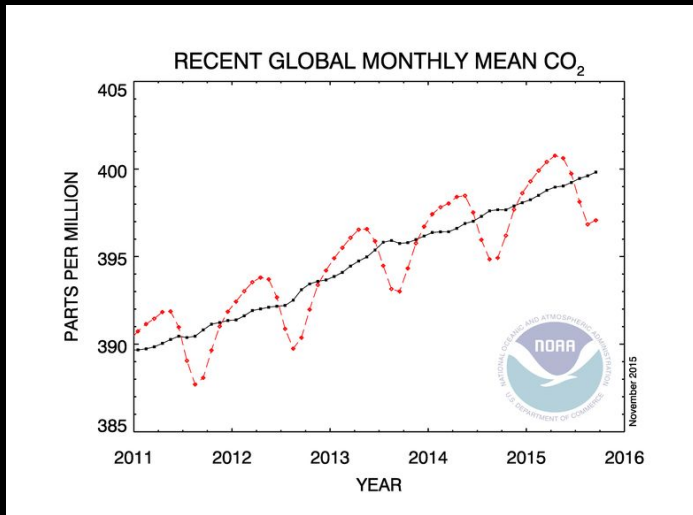


→ *Price*



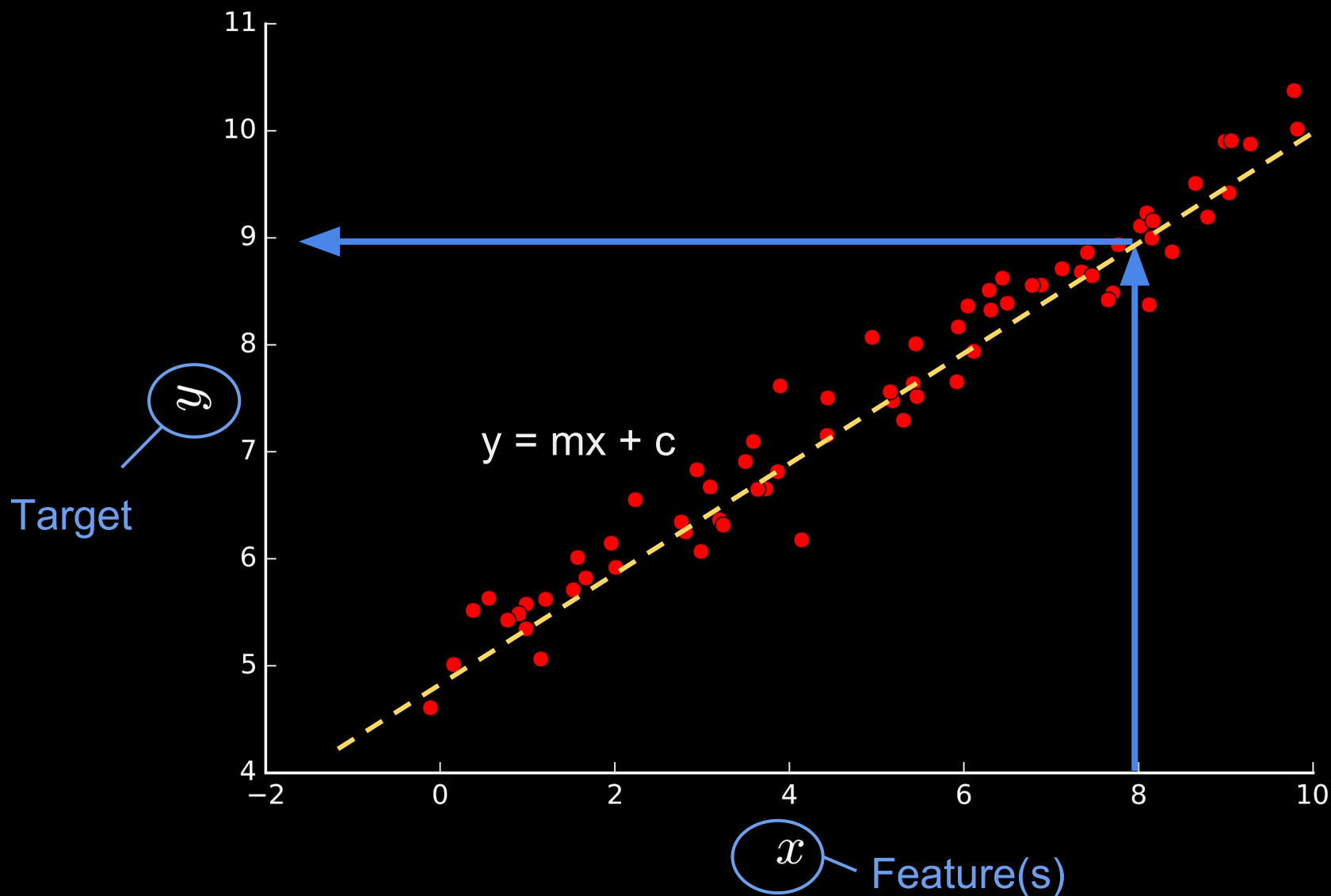
→ *Species richness*

Regression

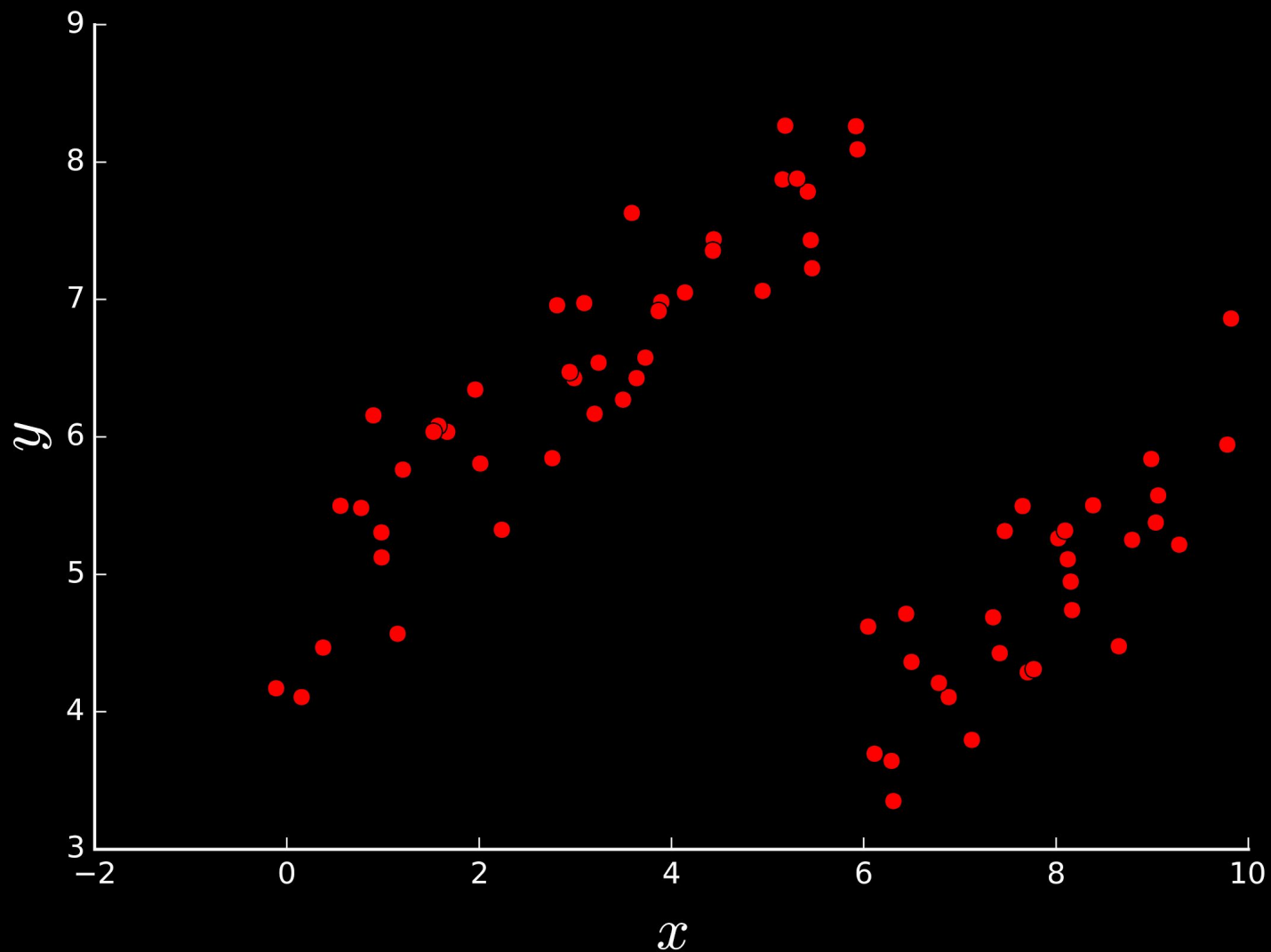


→ *CO₂ levels next decade*

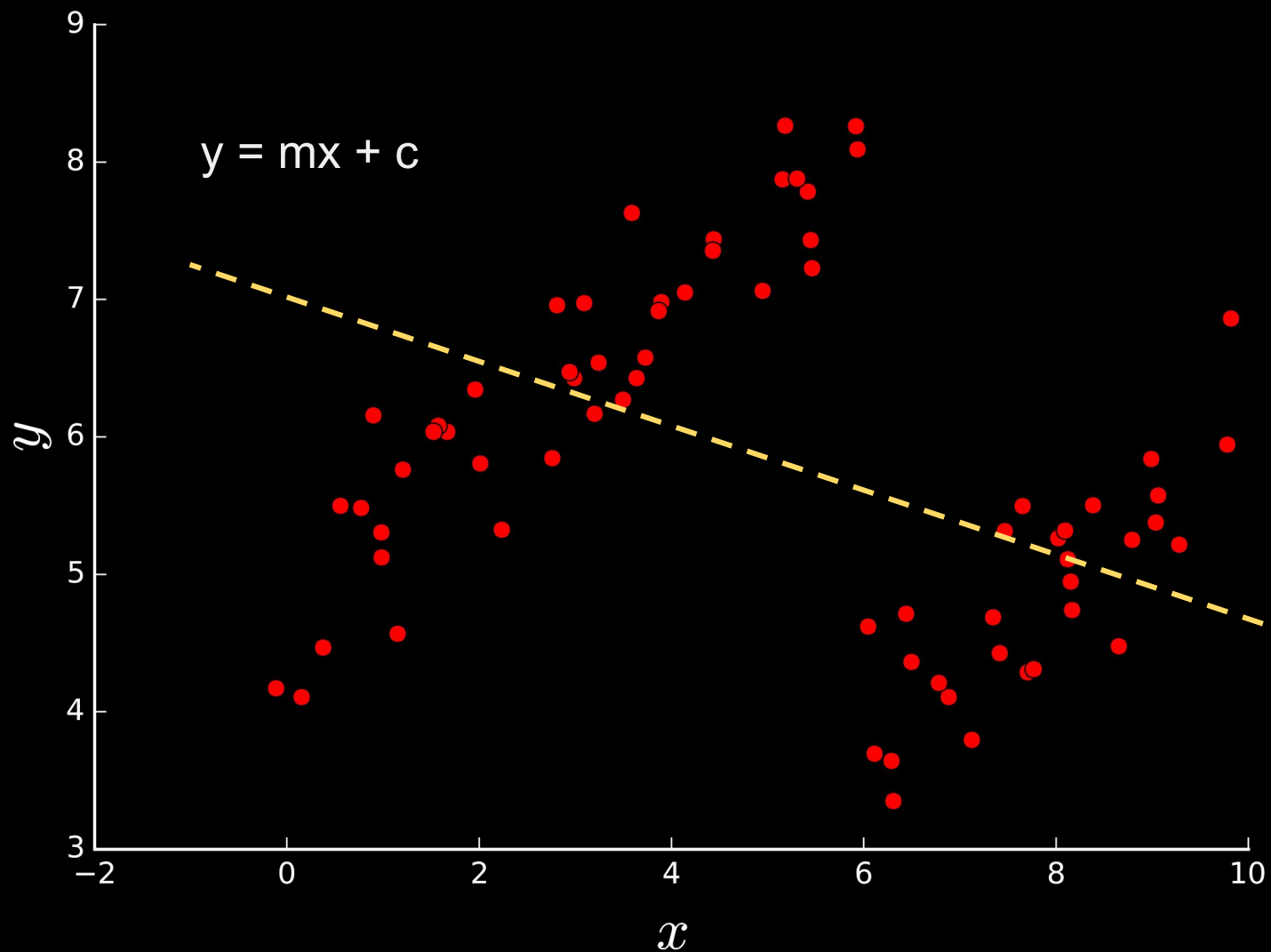
How to do regression?



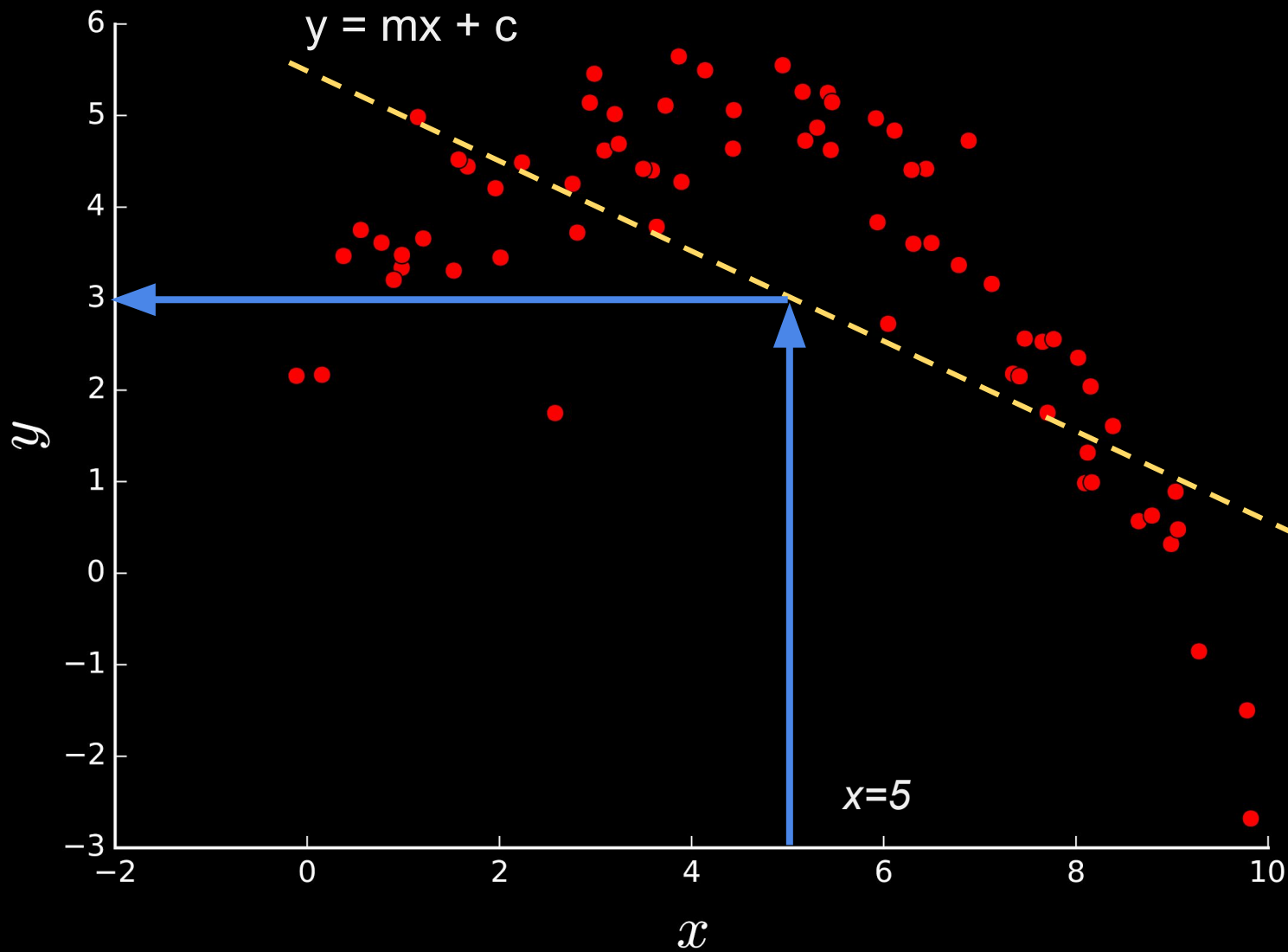
But what about non-linear data?



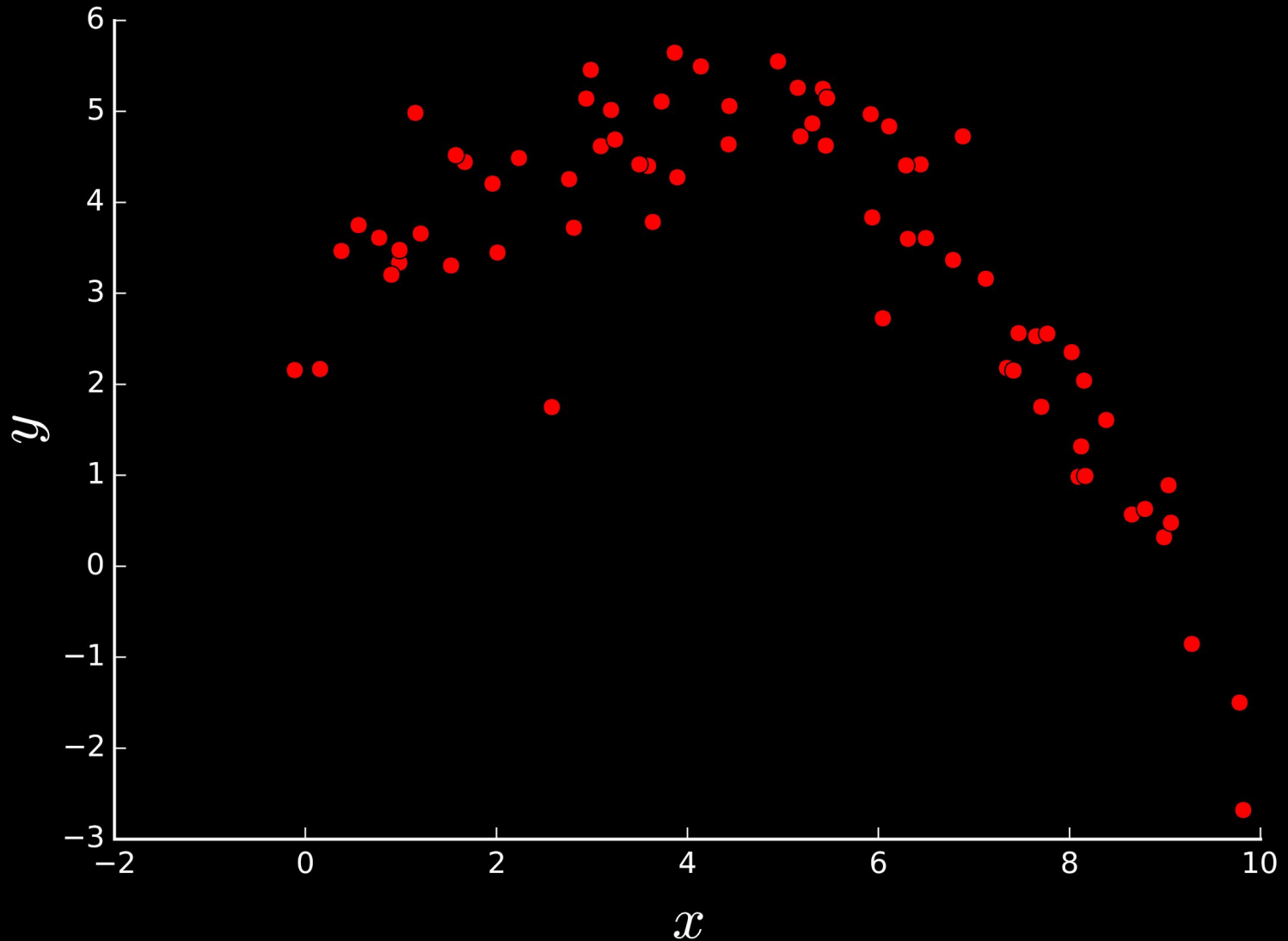
But what about non-linear data?



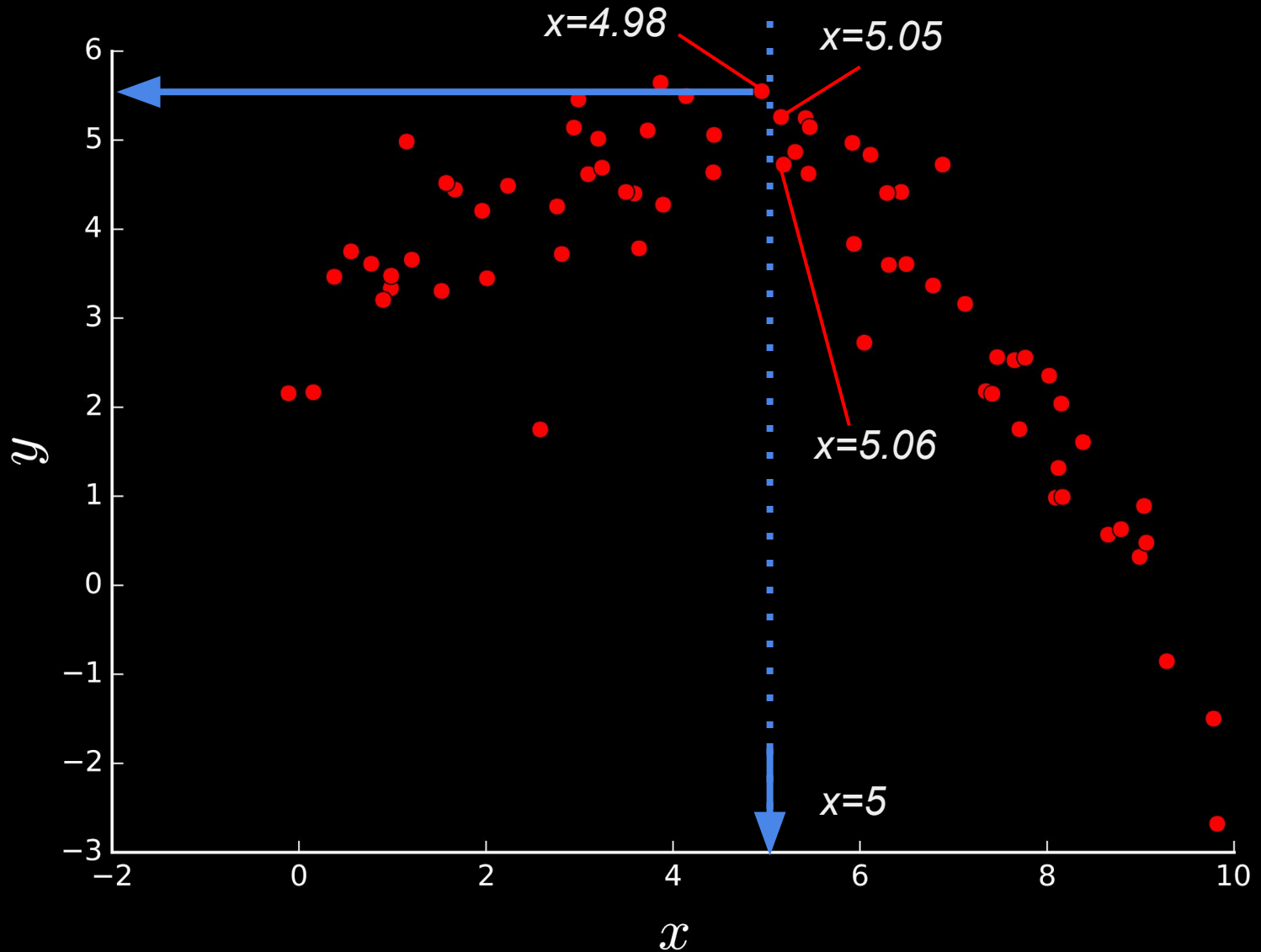
But what about non-linear data?



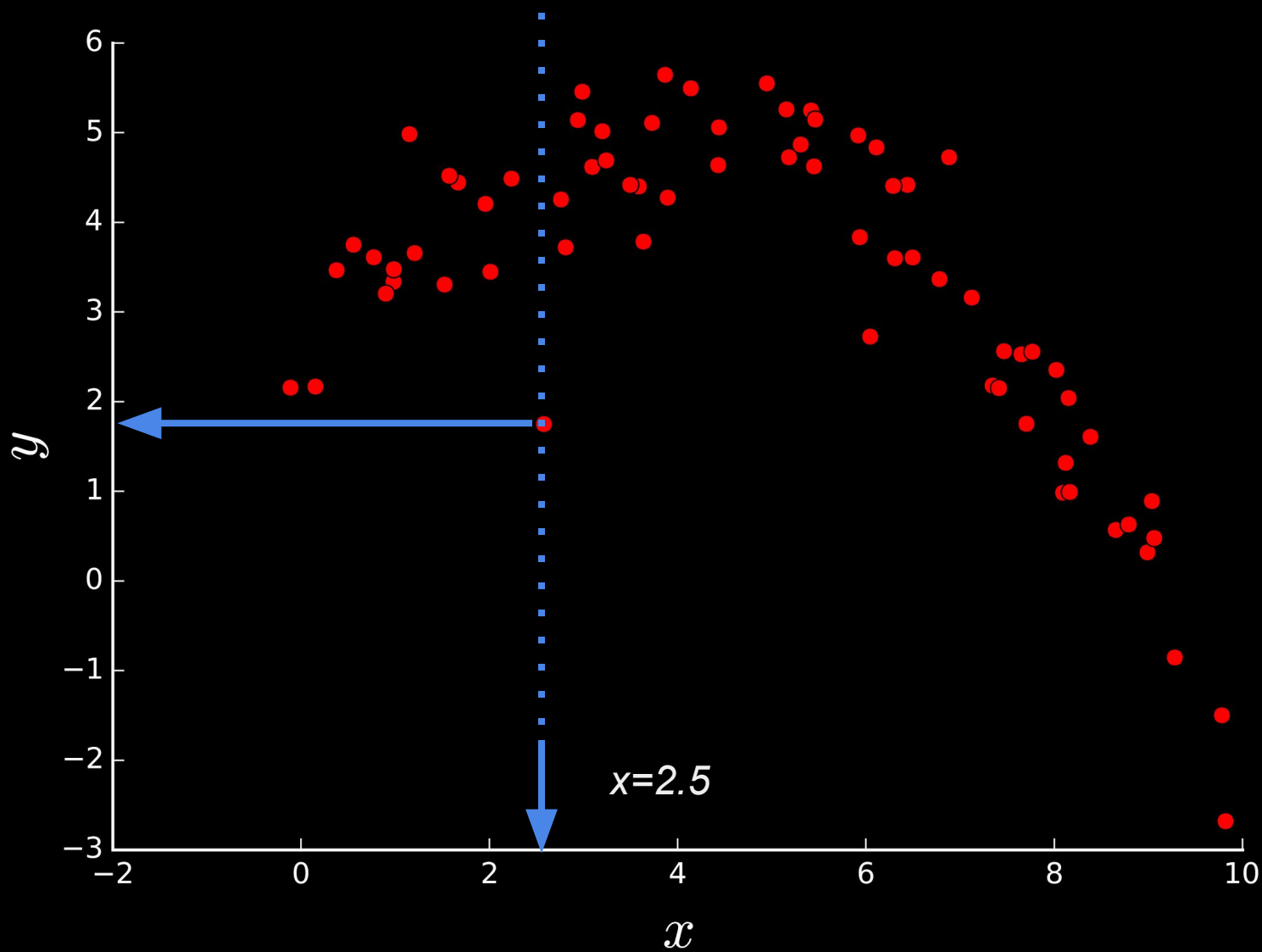
Alternative – Treat the same as classification!



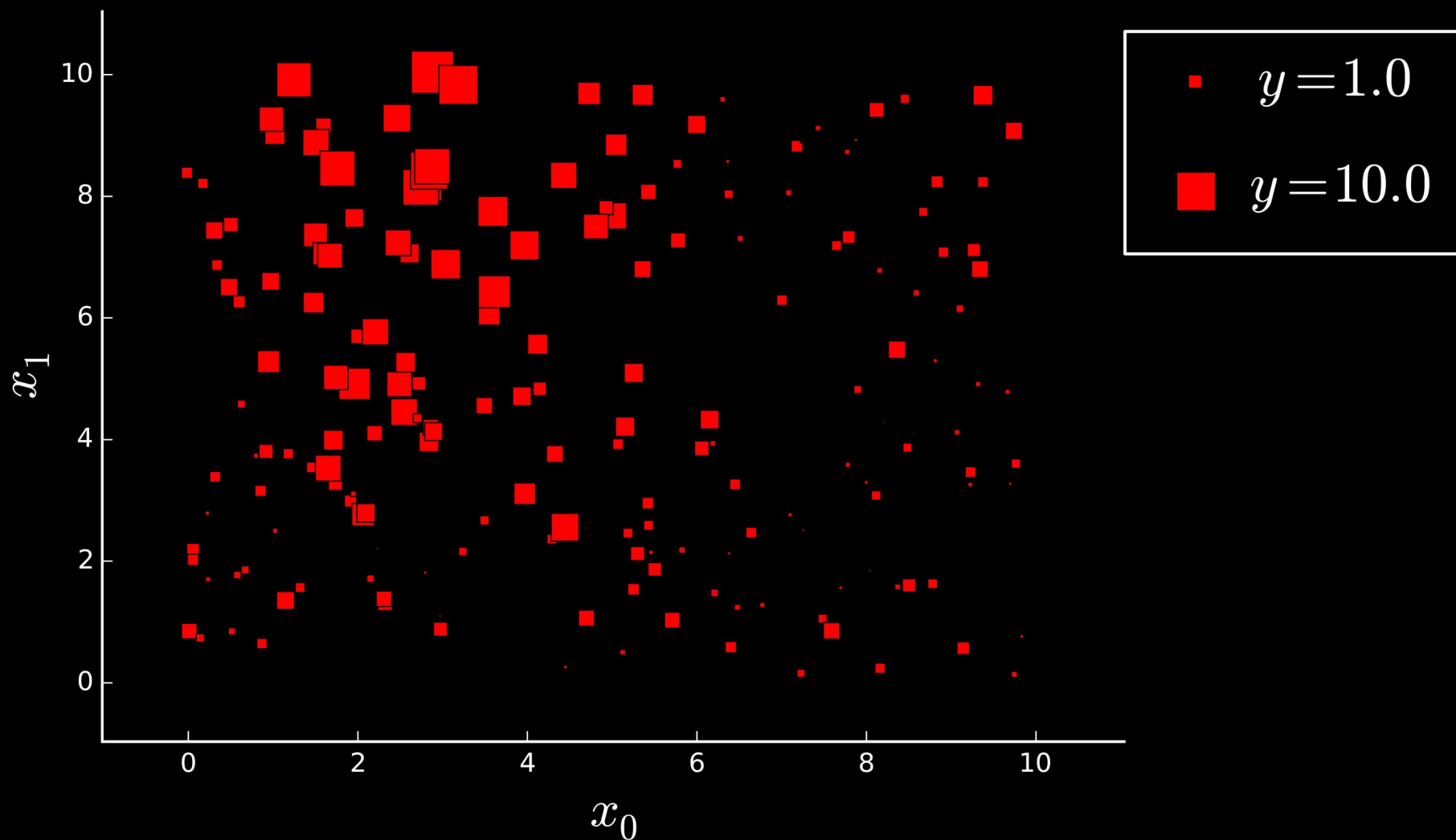
Alternative – Treat the same as classification!



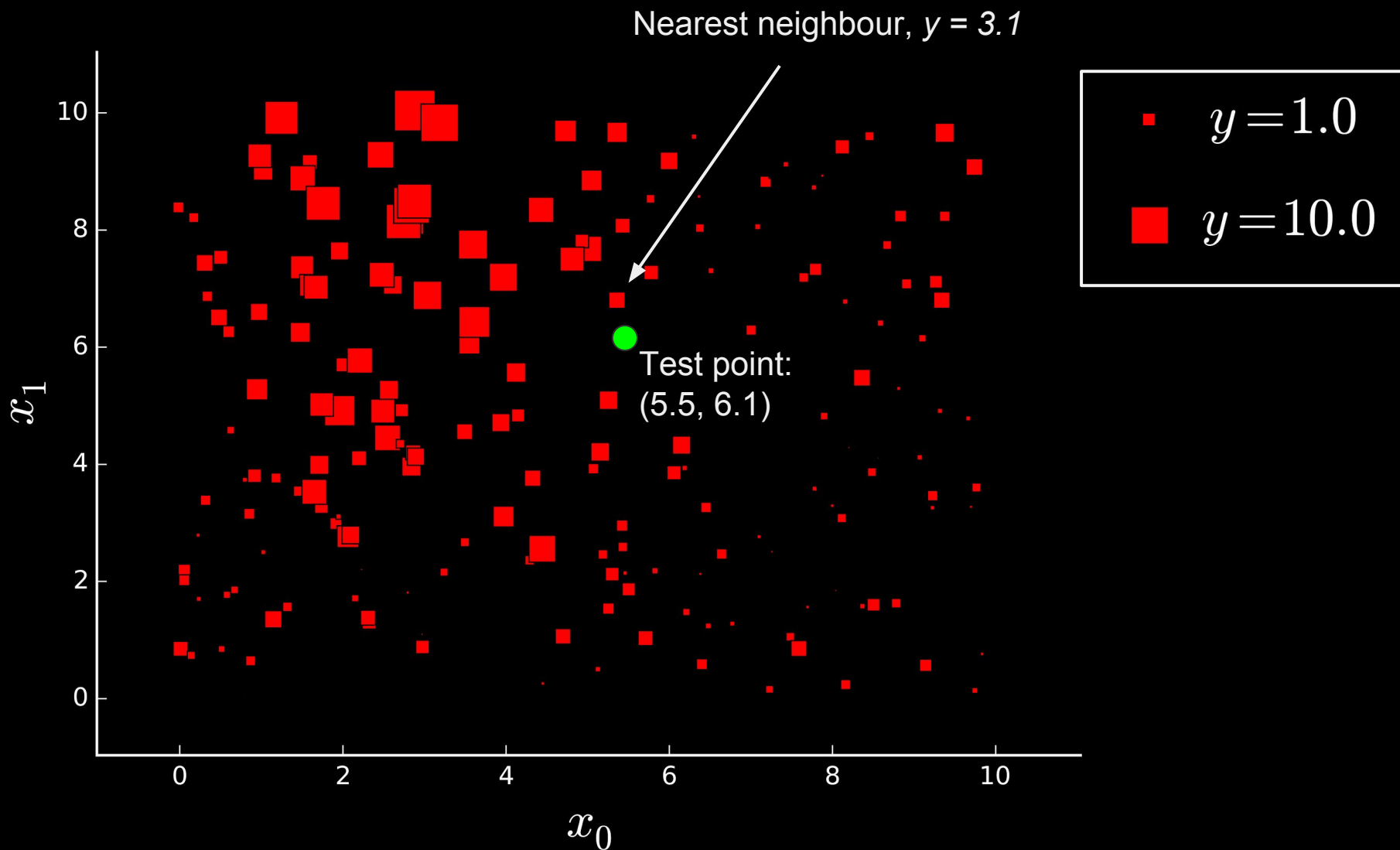
Nearest neighbours can still suffer from noise



Features can still be multidimensional



Making prediction for $x = (5.5, 6.1)$



Are we doing good regression?

- Create **train/validation split** just like in classification!
- Can't use classification accuracy as measure
- Instead, perhaps:
 - **Mean squared error**
 - **Mean absolute error**

Ground truth	Prediction 1	
5	5	error = 0
10	9	error = 1
8	8	error = 0
2	2	error = 0
5	4	error = 1
1	13	error = 12
13	12	error = 1

Practical example

4_regression.R