PSY9511: Seminar 2

The basics of regression and classification

Esten H. Leonardsen

26.02.2024

UNIVERSITETET
I OSLO

Outline

Today's lecture:

1. Recap of last lecture
2. Proposed solution for Assignment 1
3. Basics of regression and classification
4. Presentation of Assignment 2

Recap

What is statistical learning?

Recap

What is statistical learning?

- Inferentiental view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y.

Recap

What is statistical learning?

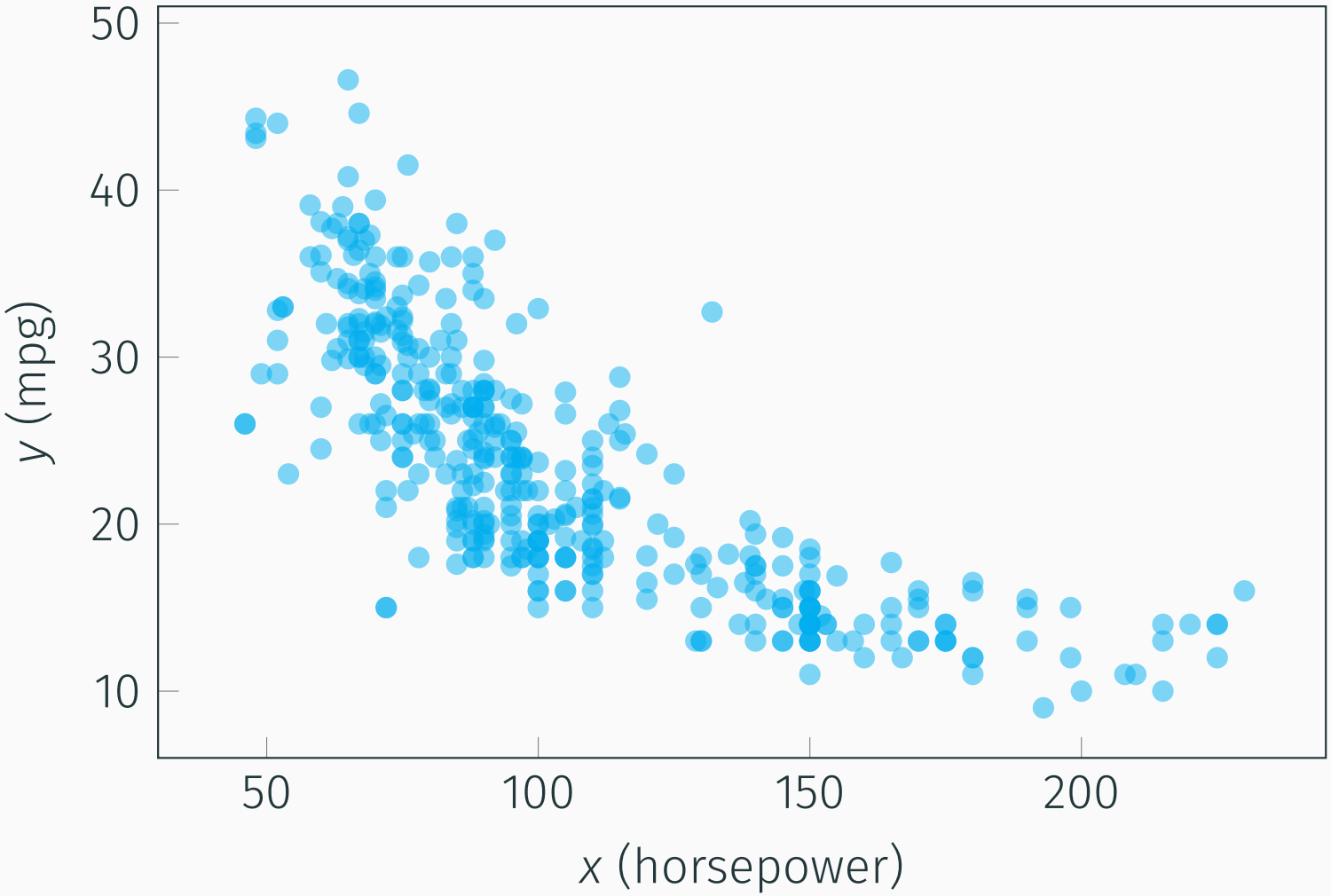
- Inferentiental view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y.
- Predictive view: Finding a function $\hat{f}(X)$ that, when given a new set of inputs X allows us to predict an output y.

Recap

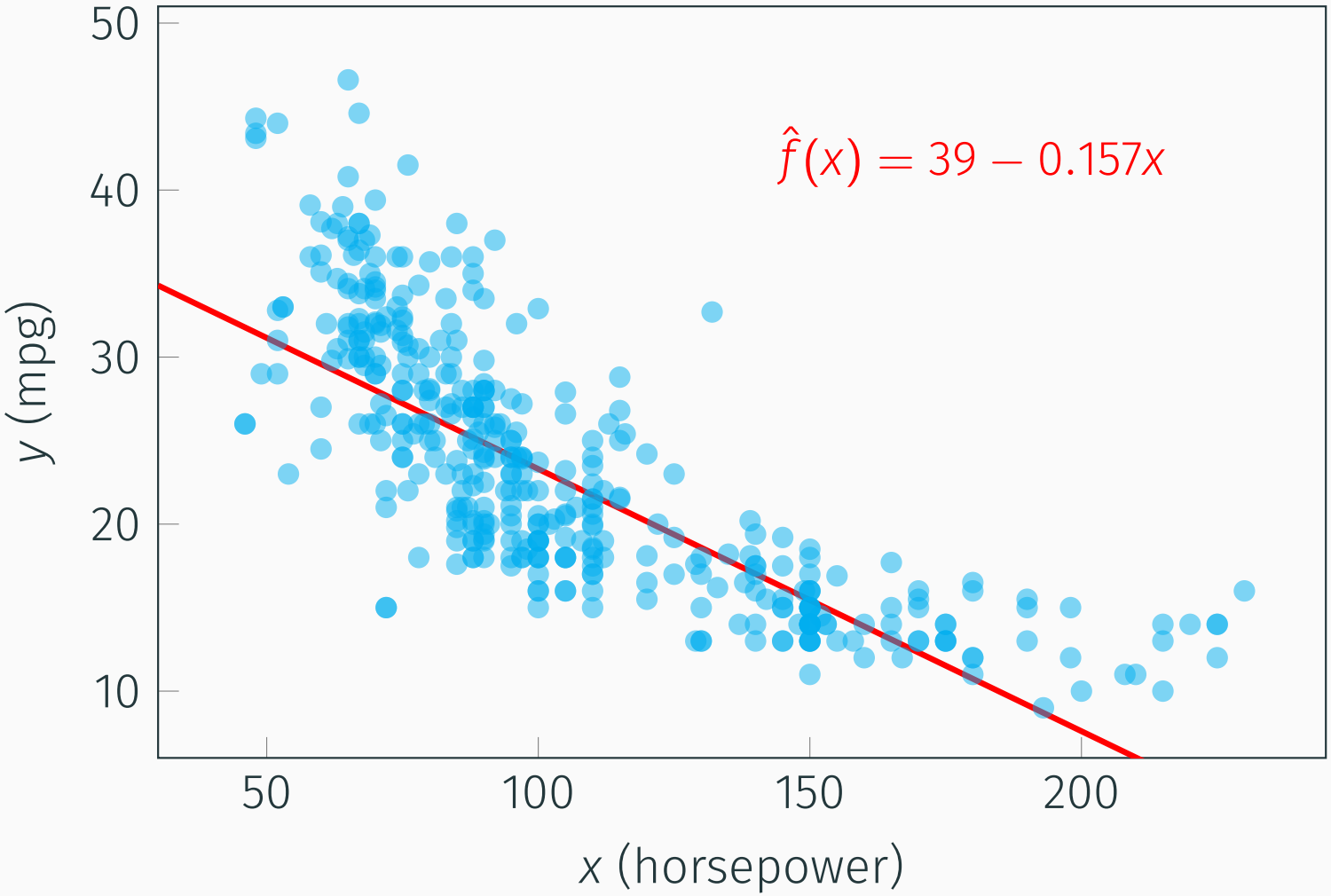
What is statistical learning?

- Inferentiental view: Finding a function $\hat{f}(X)$ that describes the relationship between some input variables X and an output variable y.
- Predictive view: Finding a function $\hat{f}(X)$ that, when given a new set of inputs X allows us to predict an output y.

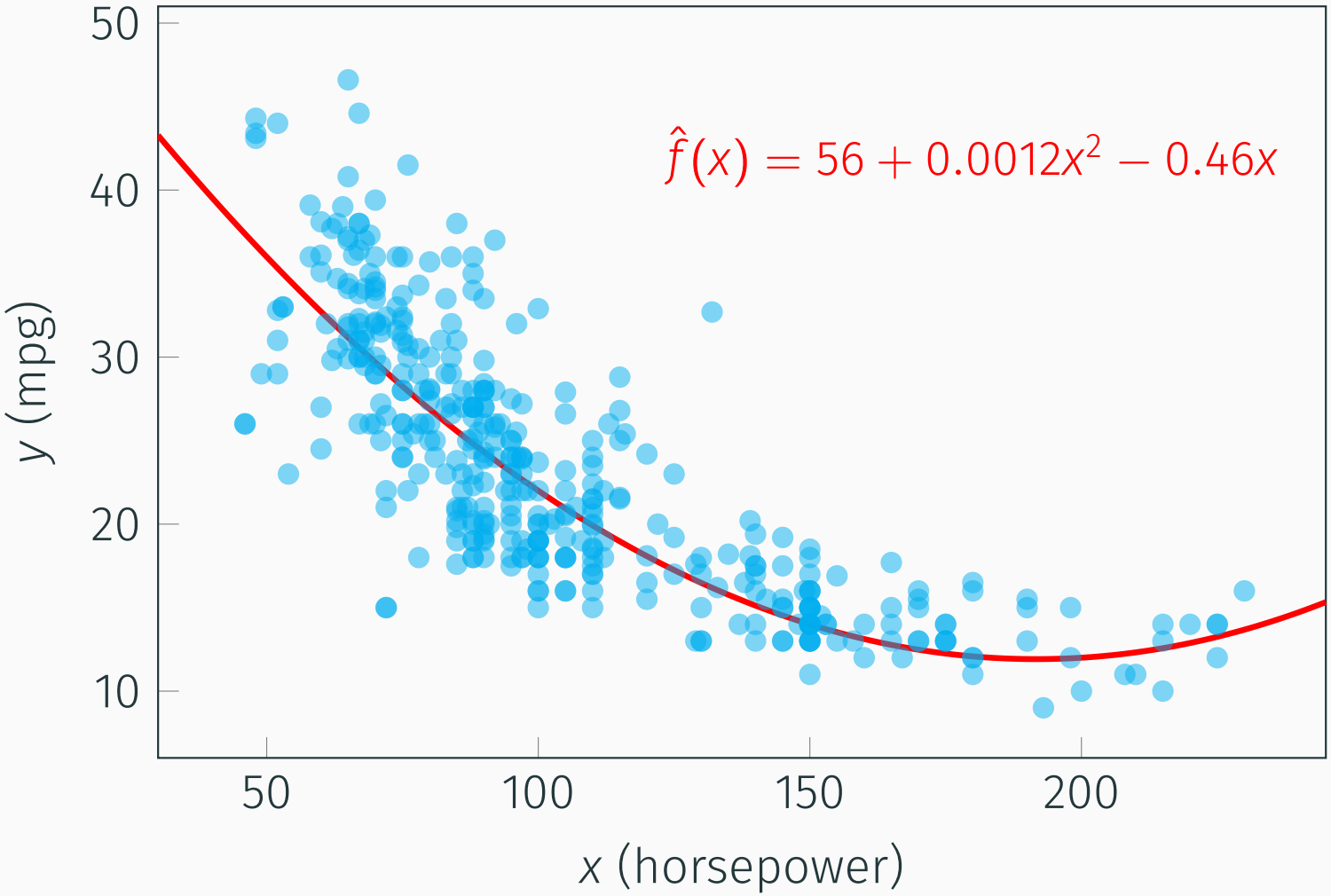
Recap



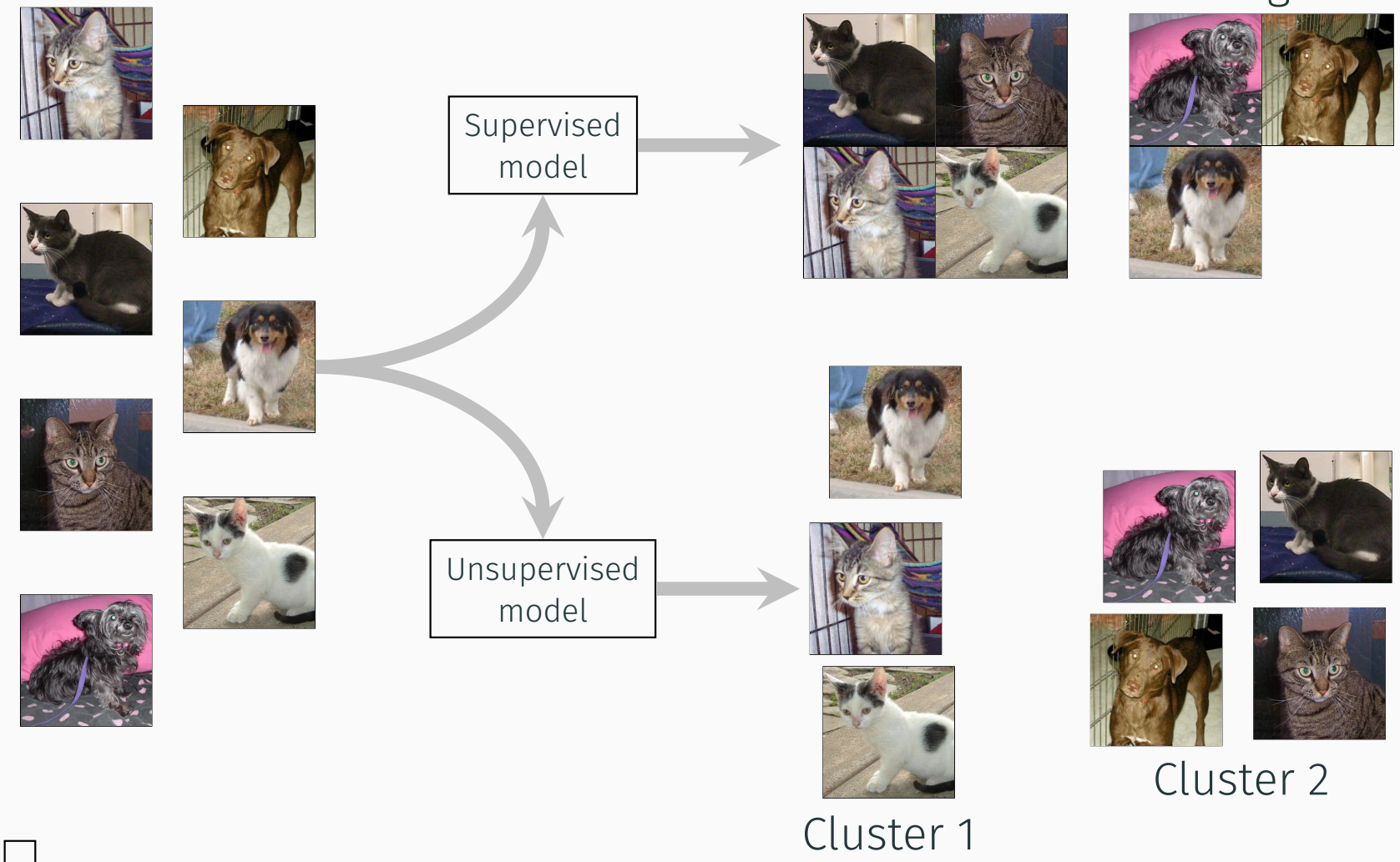
Recap



Recap


$\hat{f}(x) = 56 + 0.0012x^2 - 0.46x$
y (mpg)
x (horsepower)
50
40
30
20
10
50
100
150
200

Recap


Cats
Dogs
Supervised model
Unsupervised model
Cluster 1
Cluster 2

Recap

Regression

y
18
15
18
16
17

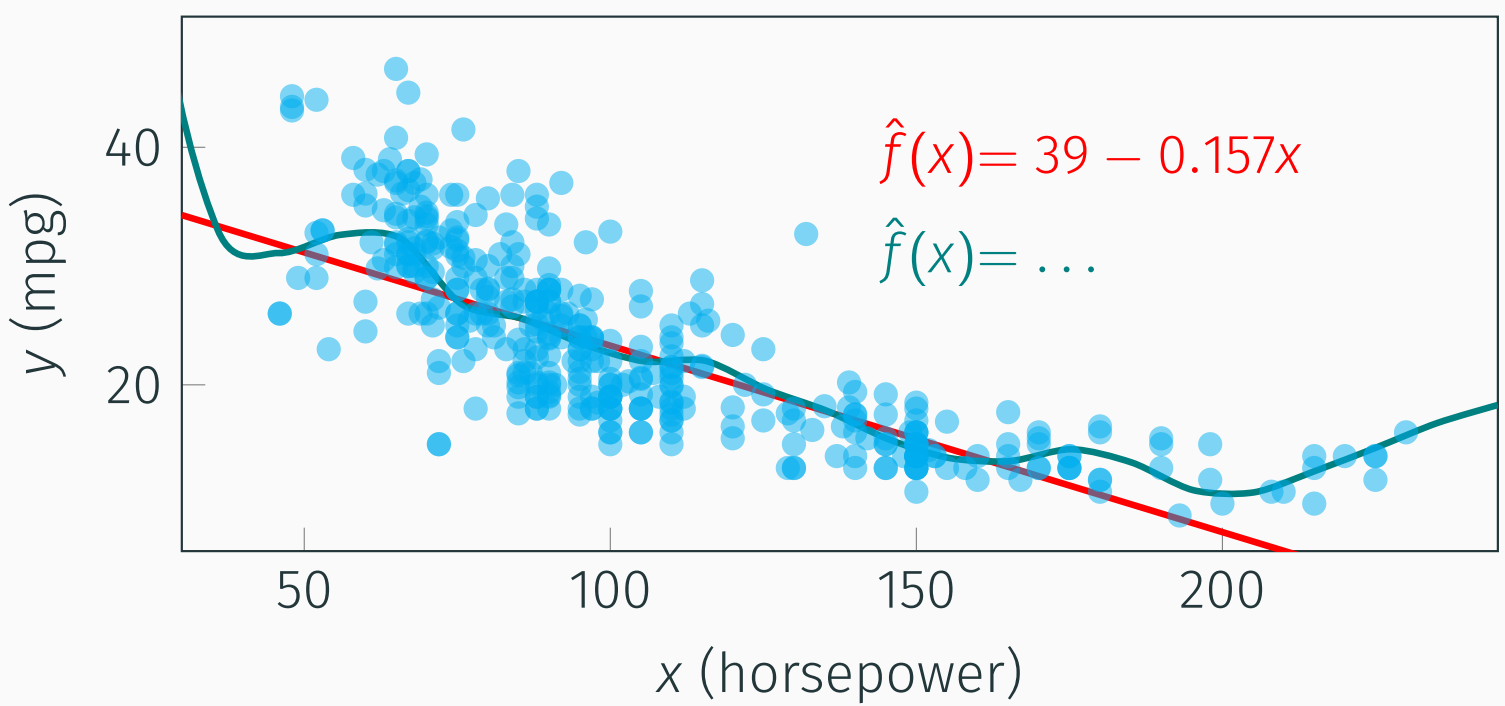
The predictive target y is a *continuous* (or *quantitative*) variable.

Classification

y
cat
cat
dog
cat
dog

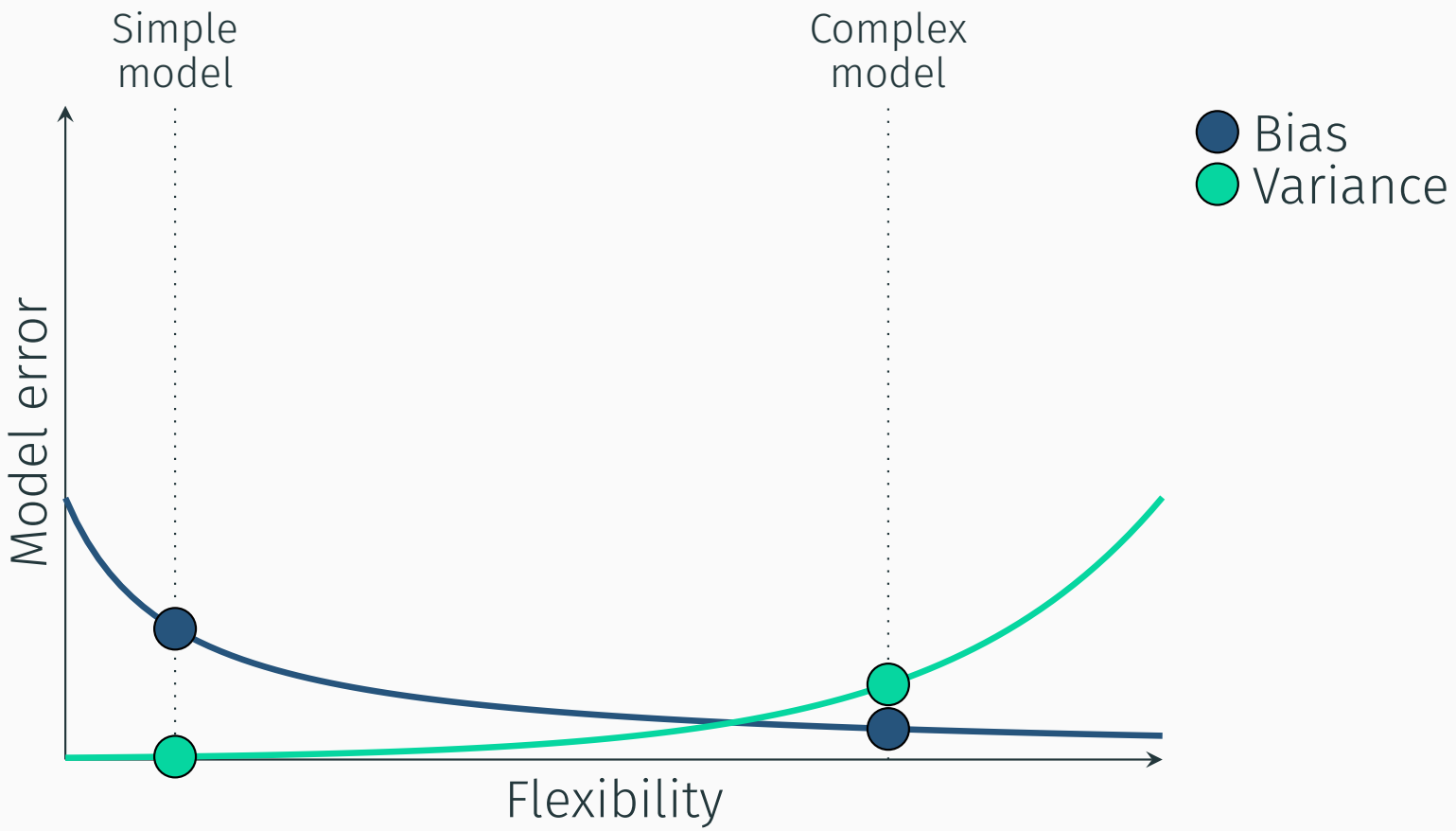
The predictive target y is a *categorical* (or *qualitative*) variable.

Recap

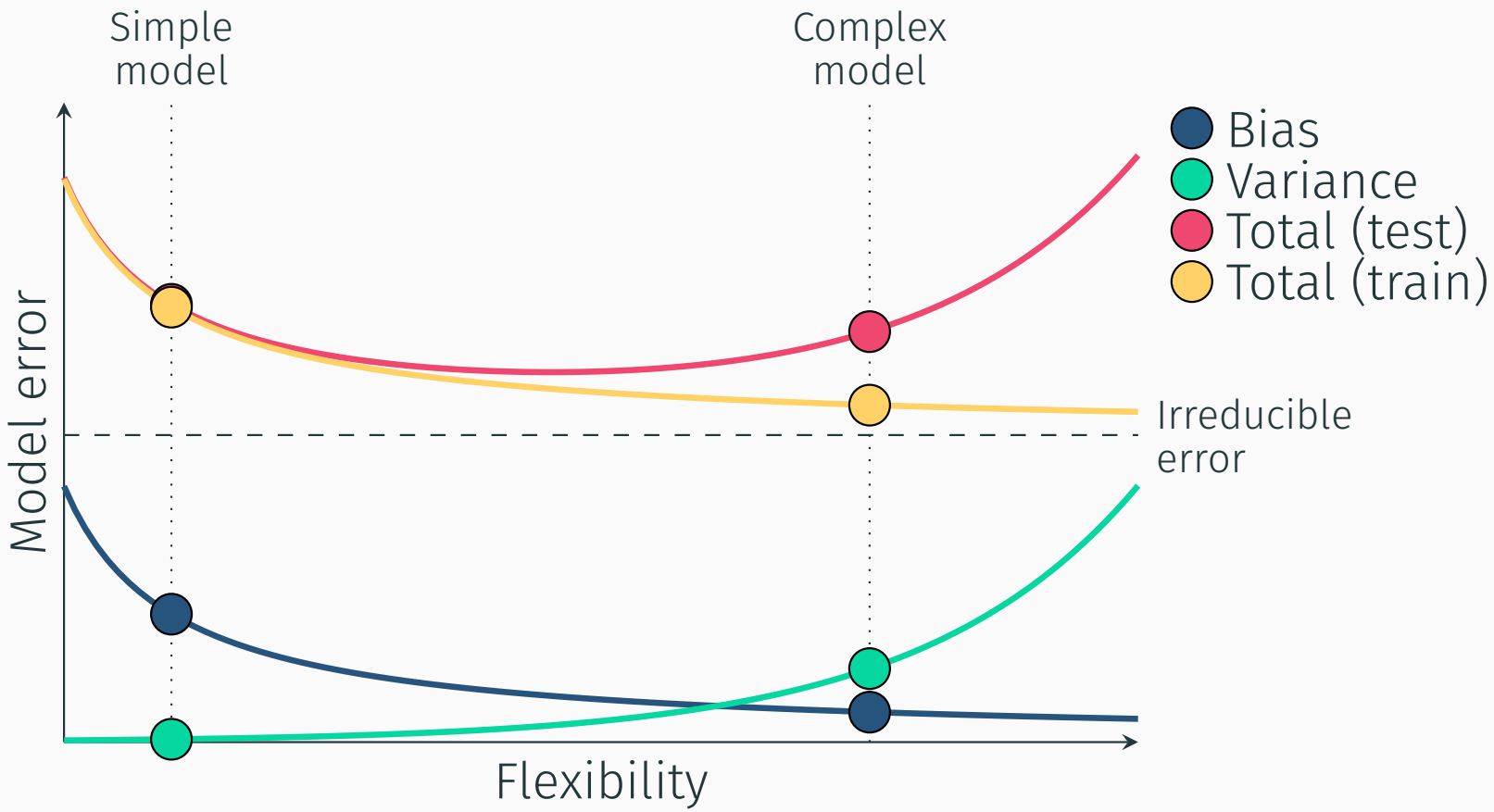


- **Parametric models** The function $\hat{f}(X)$ is relatively simple and can be described by a small number of parameters.
  - Linear regression: $\hat{f}(X) = \beta_0 + \beta_1 x$
- **Non-parametric models** The function $\hat{f}(X)$ is more complex and often relies directly on the data.

Recap


Simple model
Complex model
Bias
Variance
Model error
Flexibility

Recap


Simple model
Complex model
Model error
Flexibility
Irreducible error
Bias
Variance
Total (test)
Total (train)