FIT5145 – A very Brief Introduction to Predictive Models

Mahsa Salehi*

Faculty of Information Technology, Monash University

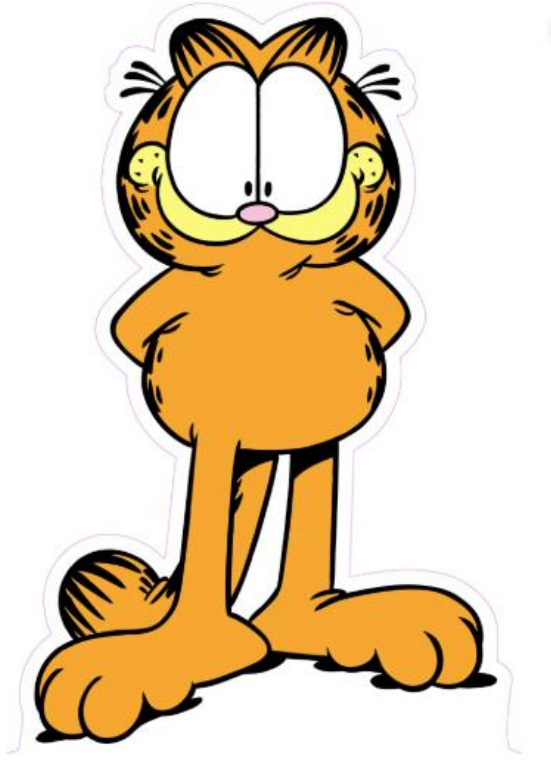
Semester 2, 2019

*with material from © Mark Carman

What is Model?

What is Model?

Can you draw a CAT..

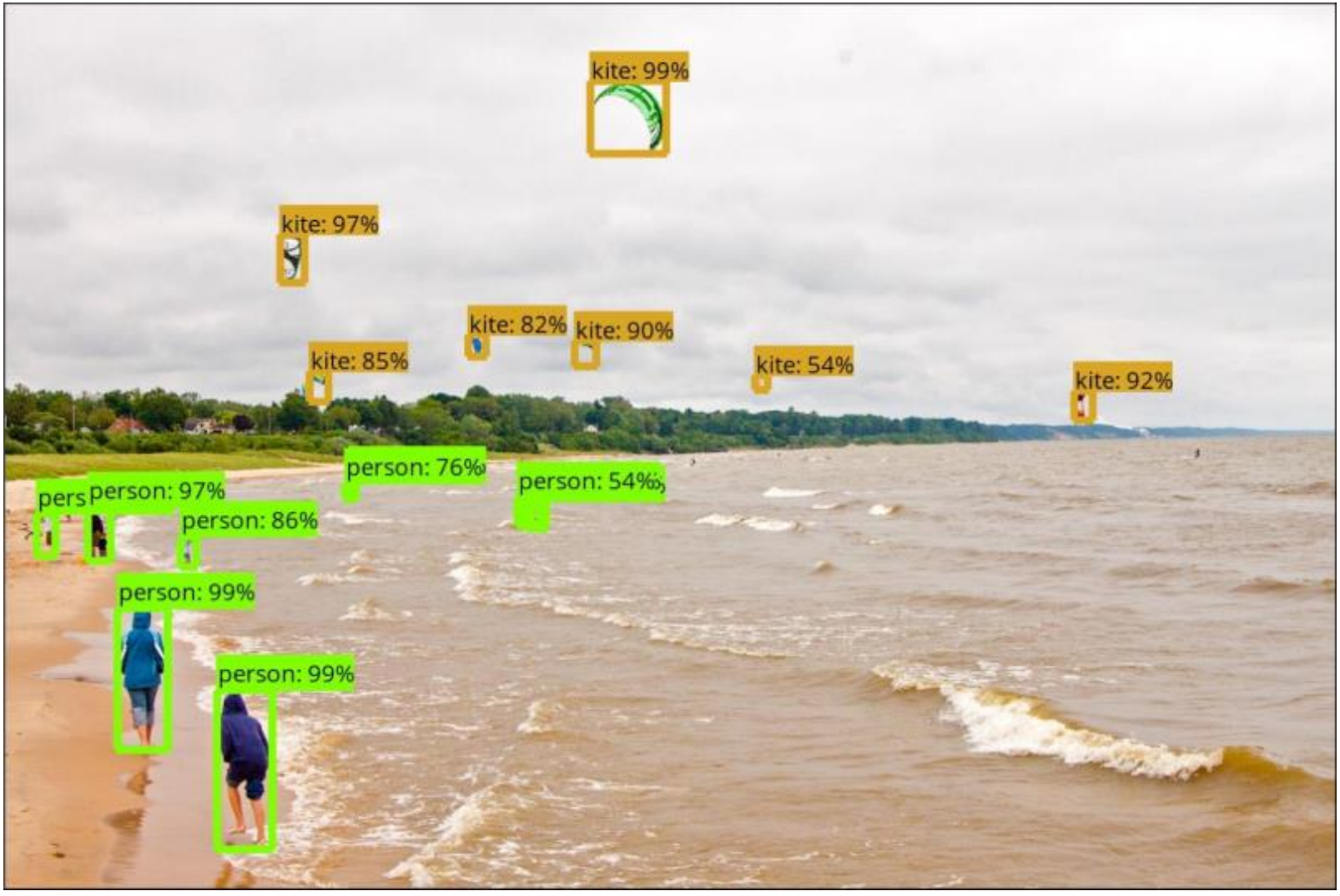


FLUX Question

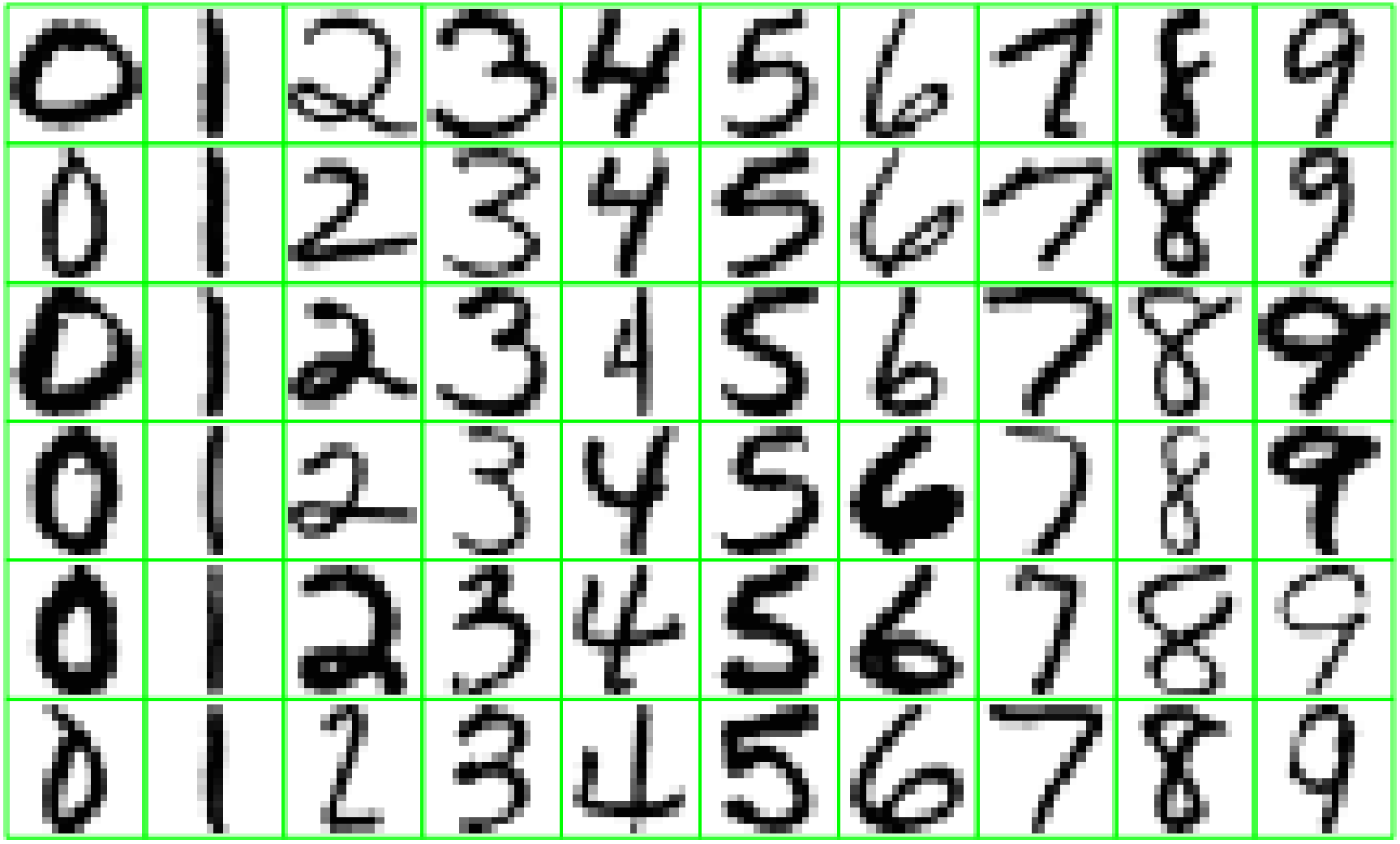
Do you think you drew a perfect model?

A. Yes
B. No
C. Not sure

What is Model?



What is Model?



FLUX Question

Which group does this horse belong to?

Group A

Group B

A brief Introduction to Predictive Models For Data Science

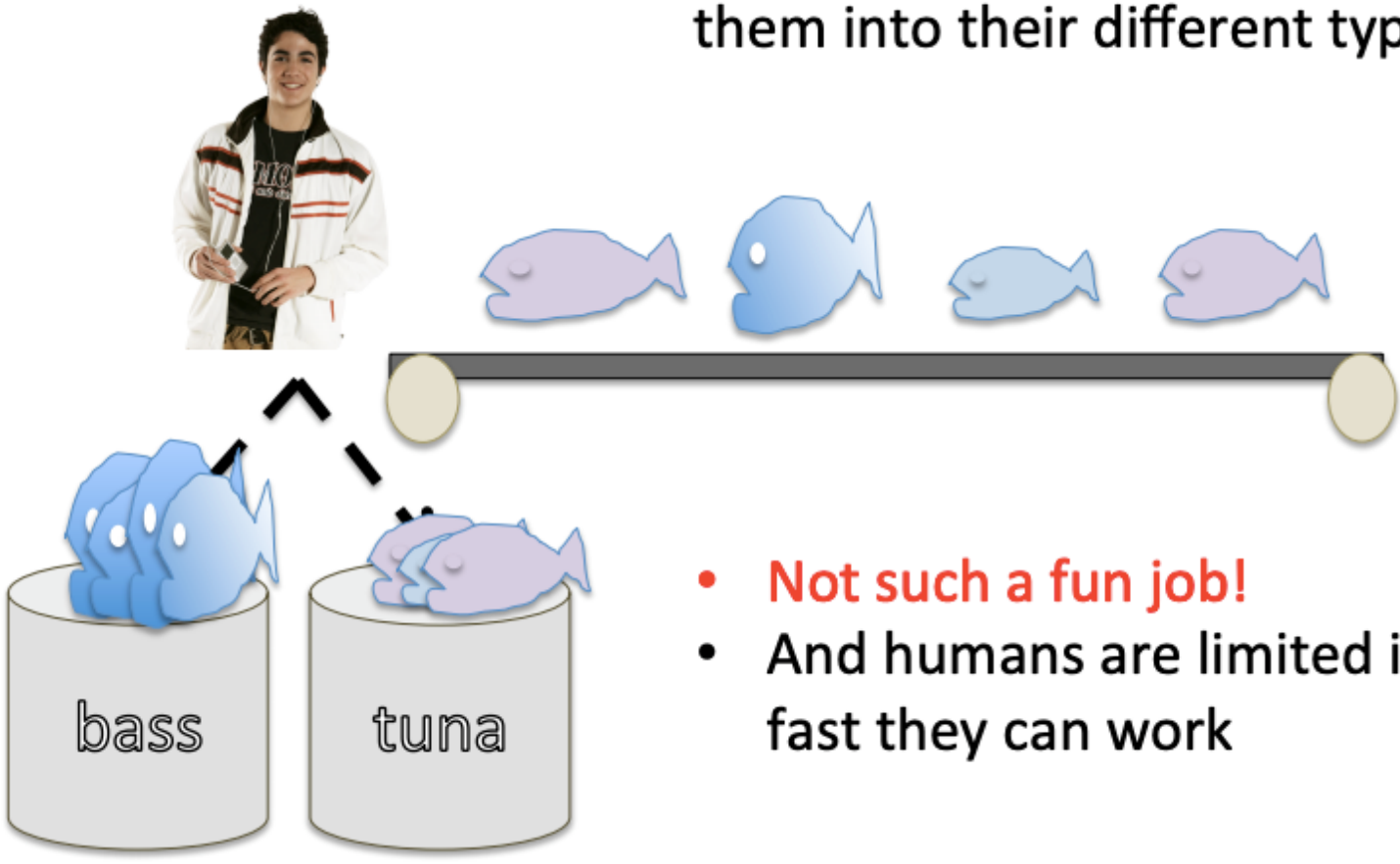
(Example from Duda & Hart, PaCern Classification & Scene Analysis, 1973)

Motivating Example

On a fishing boat, a conveyor belt loads fish and a worker separates them into their different types

Motivating Example

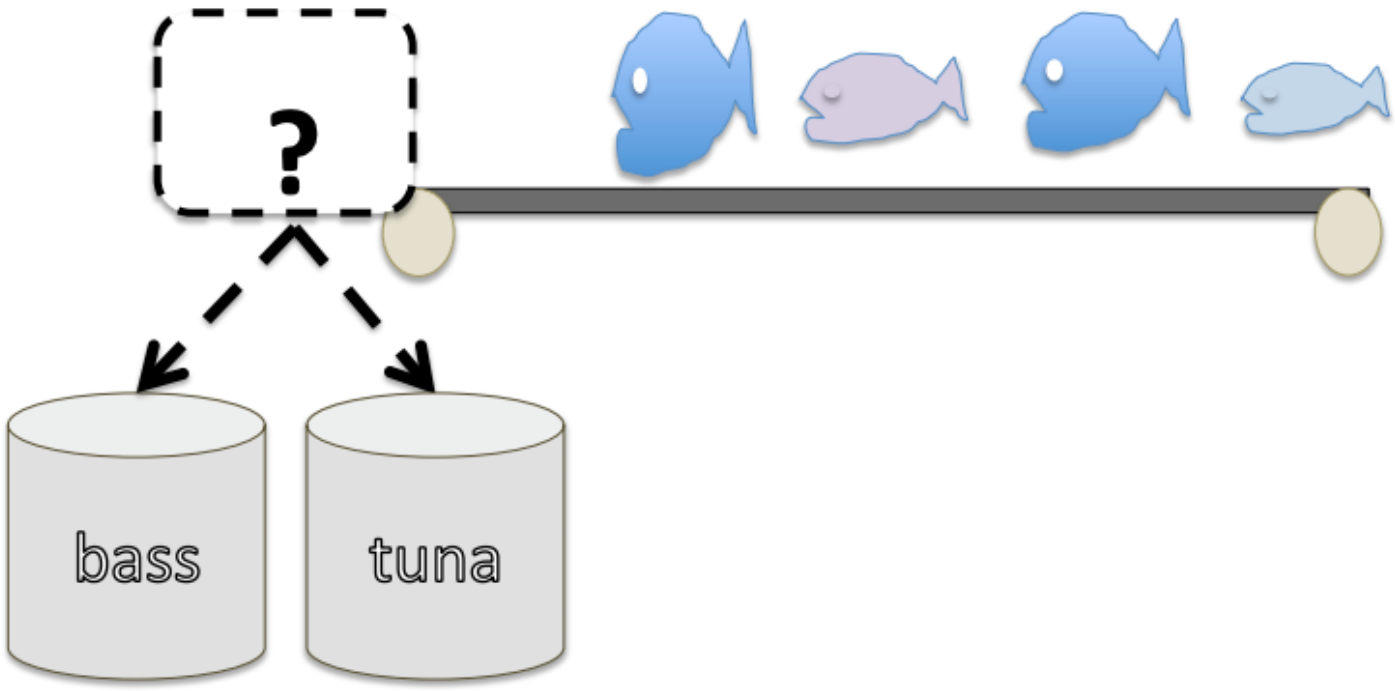
On a fishing boat, a conveyor belt loads fish and a worker separates them into their different types



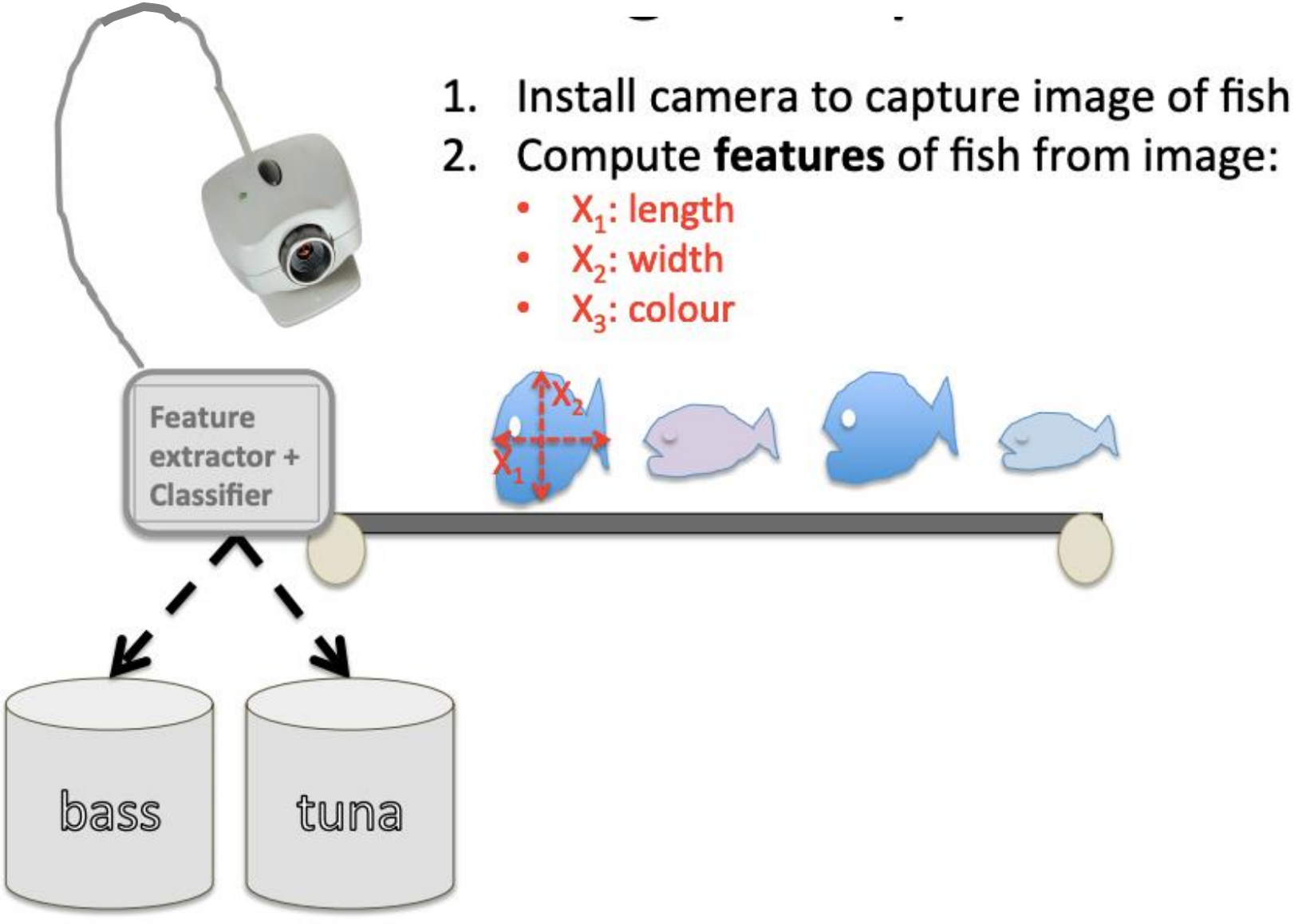
- Not such a fun job!
- And humans are limited in how fast they can work

Motivating Example

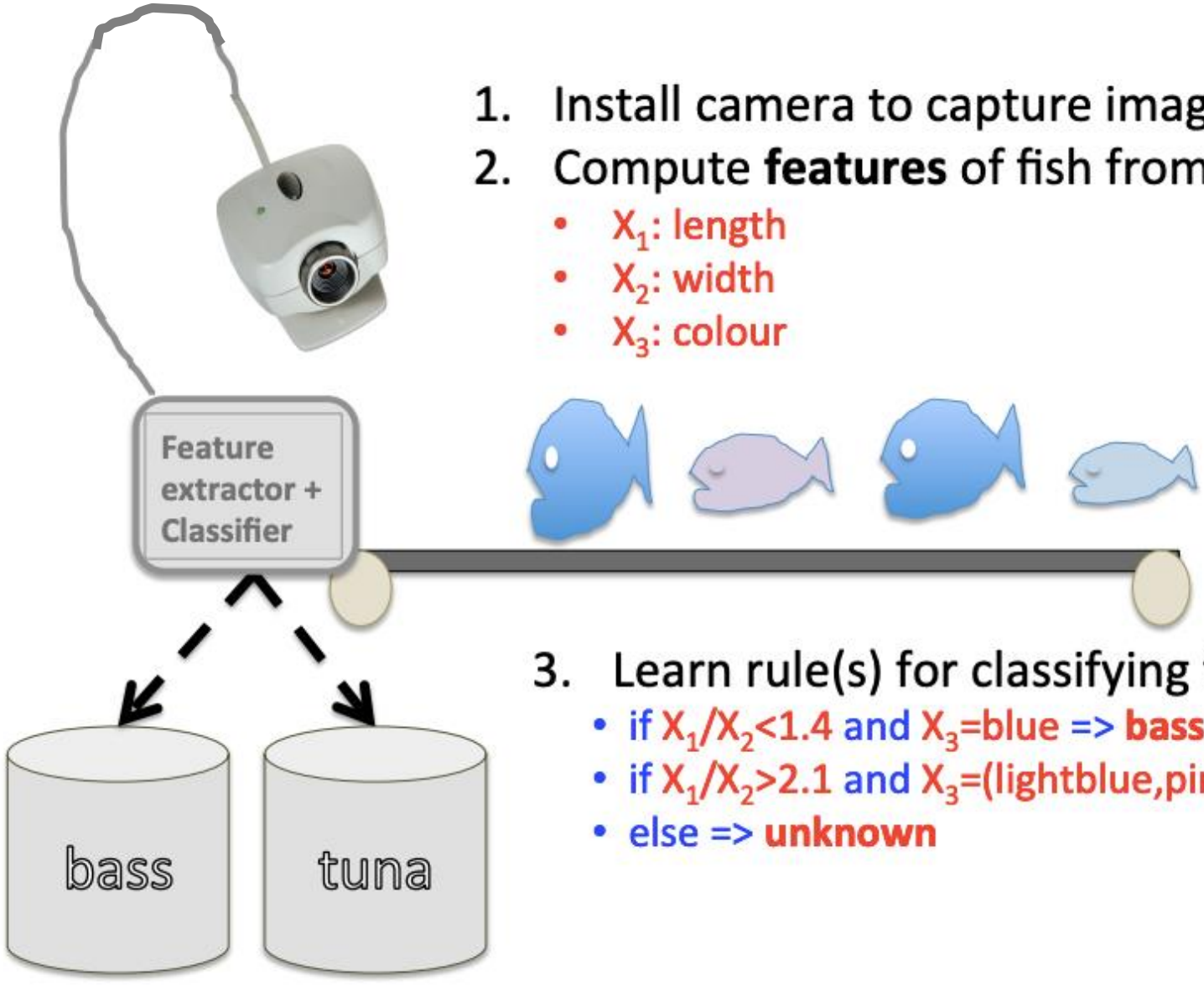
Question: Can we build a system to do the task automatically?



Motivating Example



Motivating Example



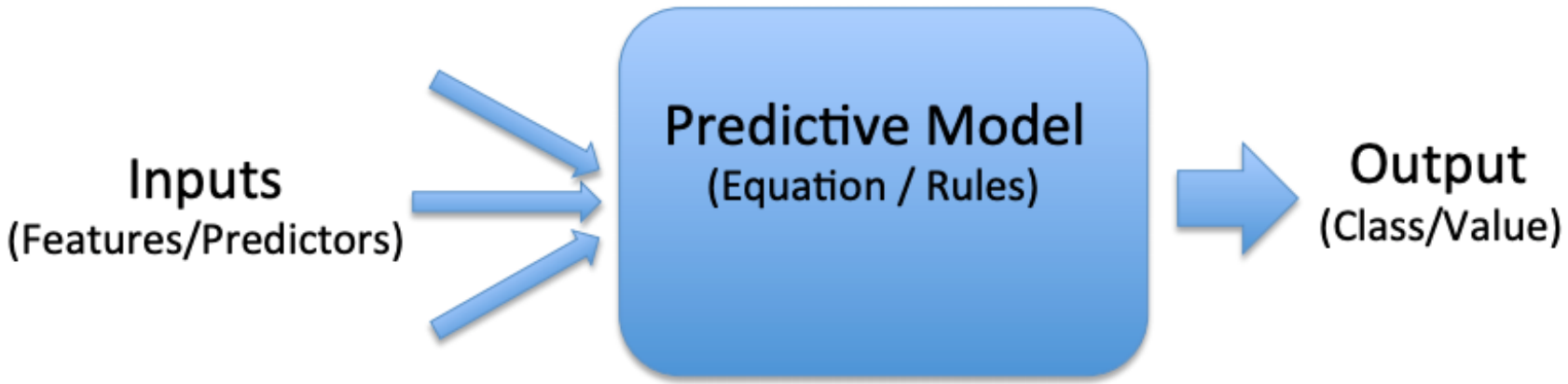
1. Install camera to capture image of fish
2. Compute **features** of fish from image:
   - X_1: length
   - X_2: width
   - X_3: colour

3. Learn rule(s) for classifying fish
   - if $X_1/X_2<1.4$ and X_3=blue => **bass**
   - if $X_1/X_2>2.1$ and X_3=(lightblue,pink) => **tuna**
   - else => **unknown**

Predictive Models

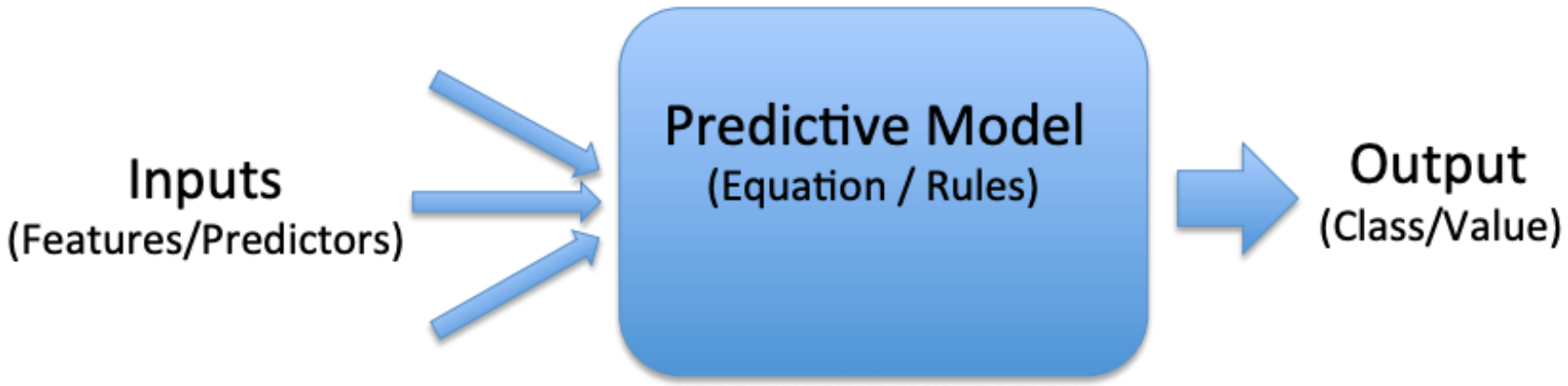
A predictive model is any model that makes a prediction

- Usually based on a set of features describing an object.
- The prediction could be:
  - A binary outcome (spam, not-spam)
  - Categorical (bass, tuna, other)
  - A real value (the age of the fish)
  - A vector of real values (probability of bass, tuna)
  - Etc.



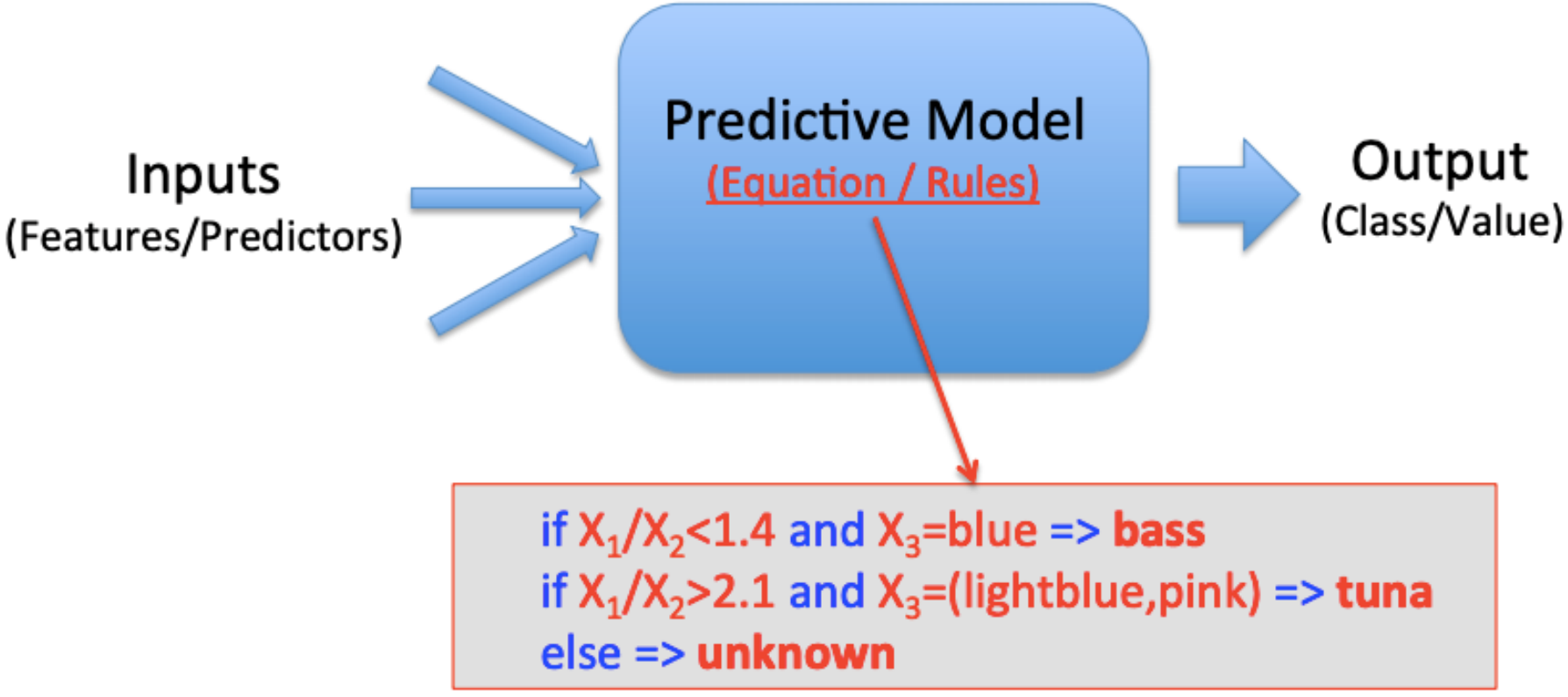
Predictive Models

- If the predicted value is binary/categorical we usually refer to the model as a classifier
- If it predicts real values we refer to it as regression
- Although there are many other types of models (e.g. ranking, translation, etc.)

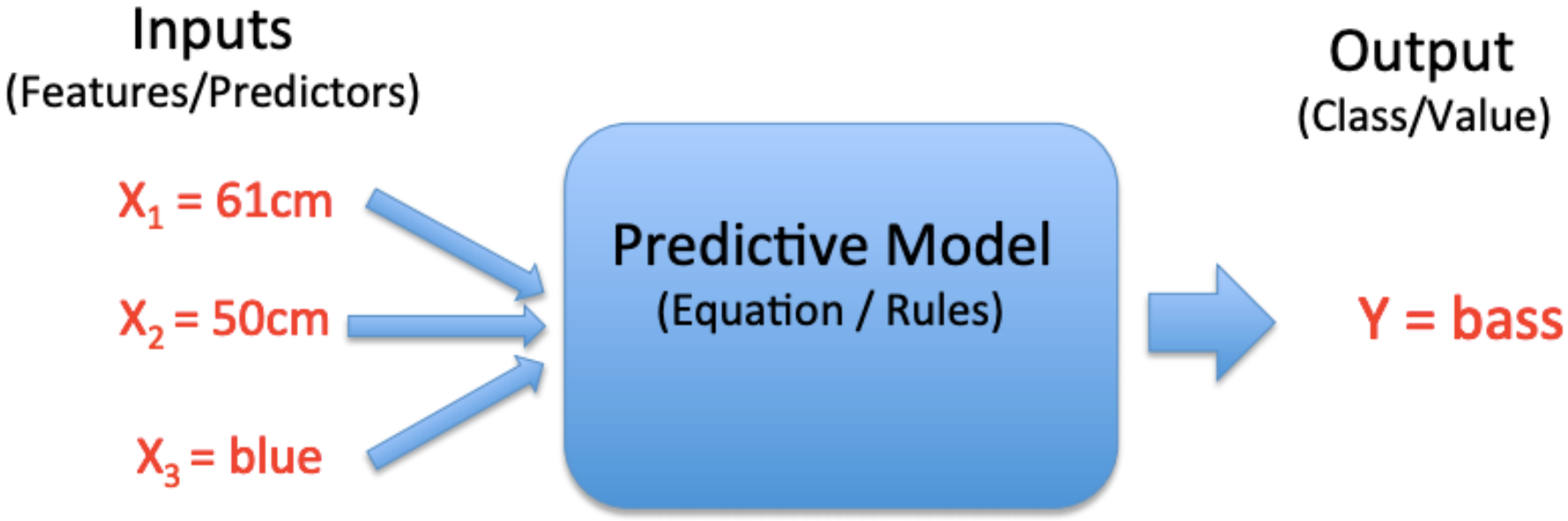


Predictive Models

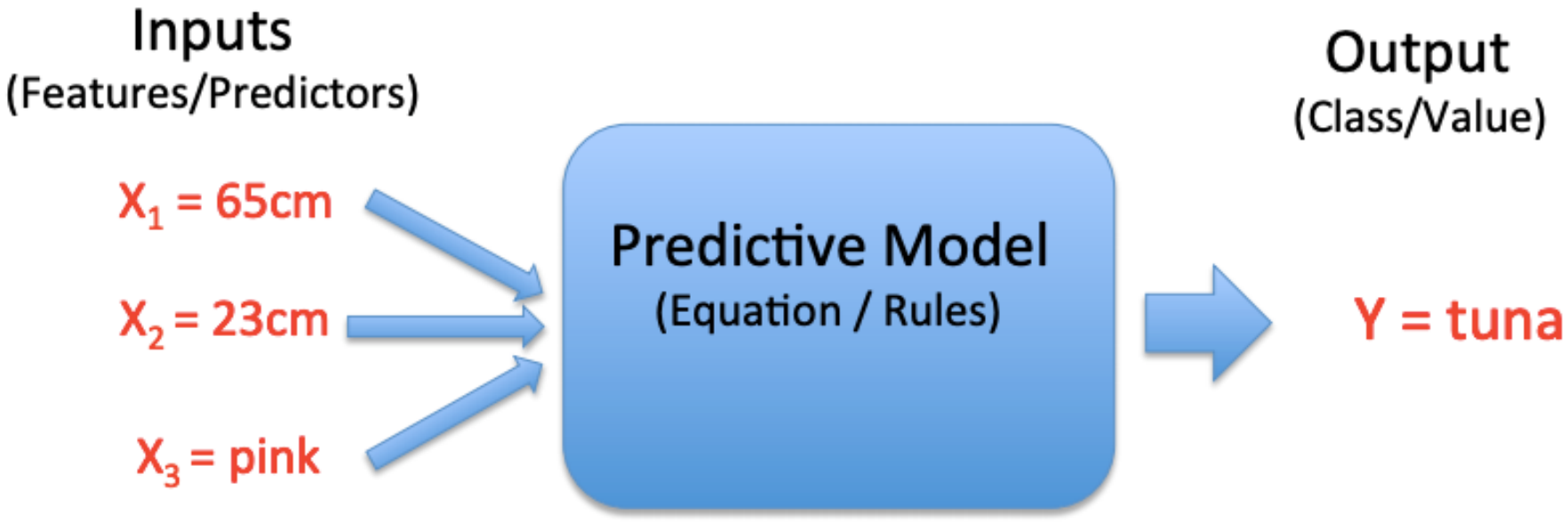
The predictive model uses equations/rules to map the input features to output values



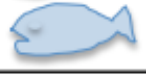
Predictive Models



Predictive Models

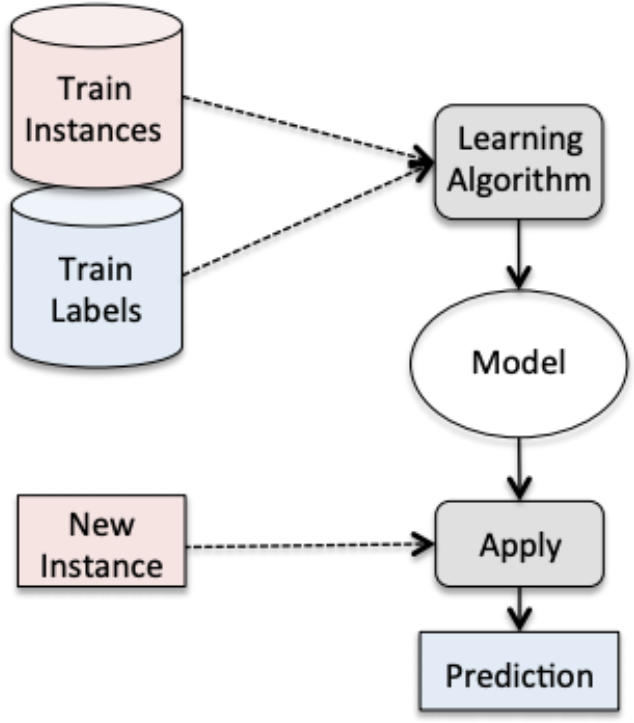


Models are learnt from Examples

Instance	X1 = length	X2 = width	X3 = colour	Y = class
	55	51	blue	bass
	65	23	pink	tuna
	67	54	blue	bass
	54	20	light-blue	tuna
	62	26	pink	tuna
	44	62	blue	bass
	47	55	light-blue	bass
	73	31	pink	tuna
	54	48	light-blue	bass
	57	23	light-blue	tuna

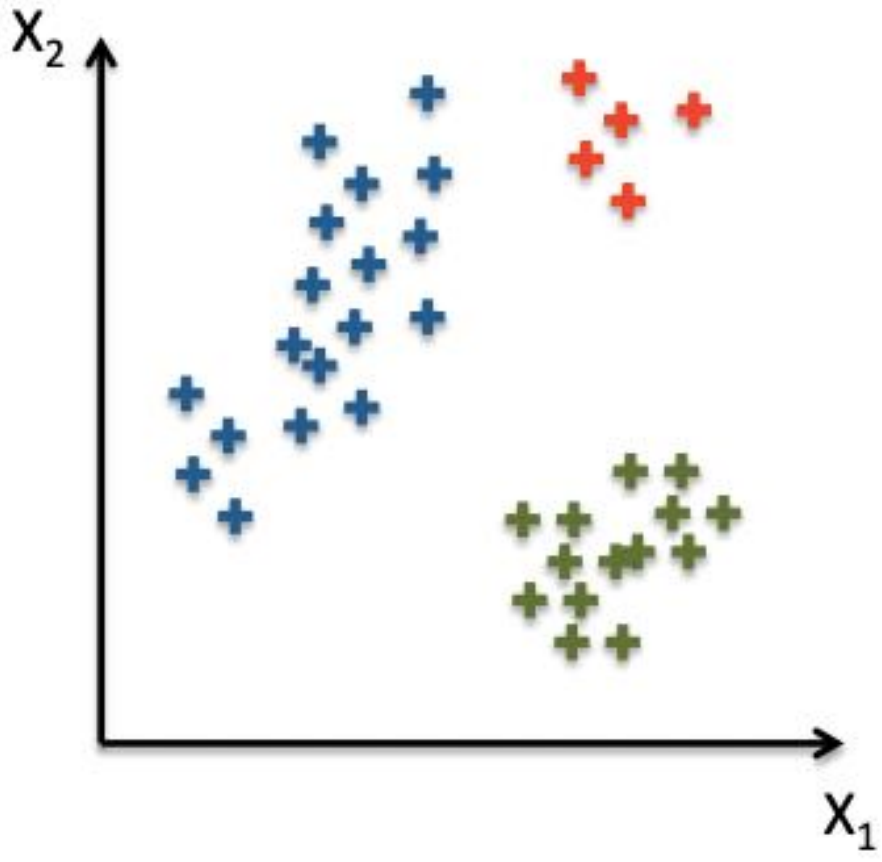
Training a Model

Predictive models are learnt from training data and then applied to make predictions on new instances



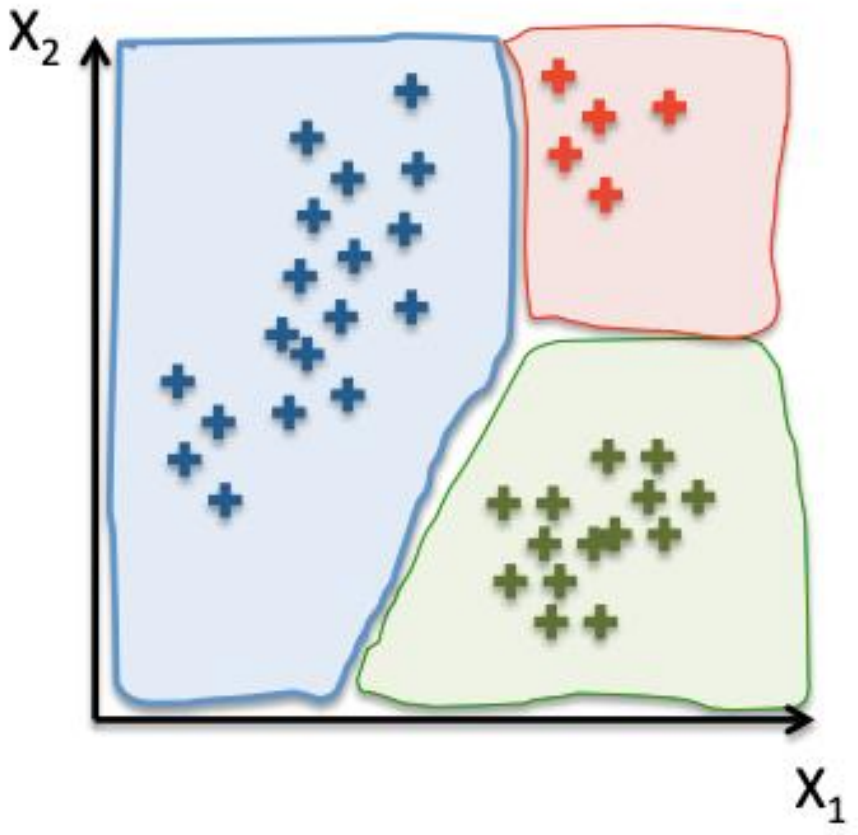
How are models learnt?

- Each training instance (fish in our case) is just a point in some feature space
- Here the colour denotes the class
  - (blue = bass, green = tuna, red = unknown)



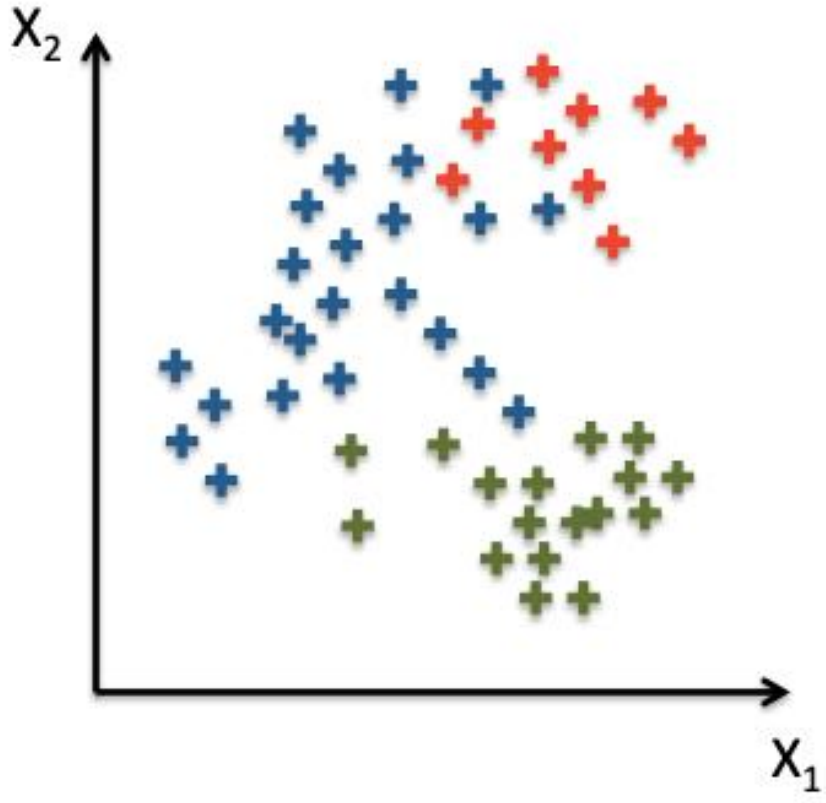
How are models learnt?

- Many (classification) learning algorithms work by dividing the feature space into regions of the same type



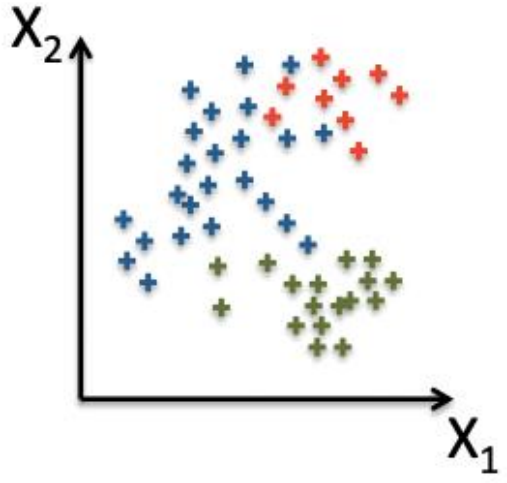
In Practice

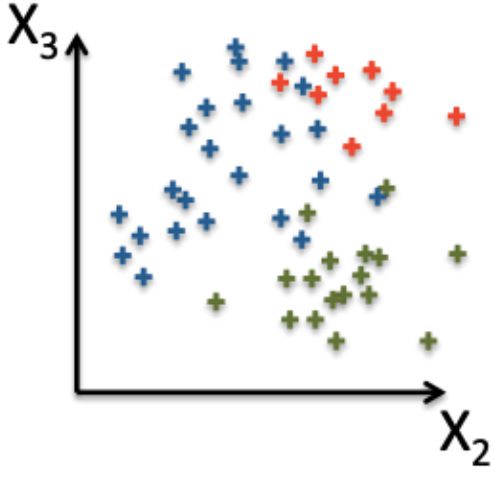
- In practice, the data is usually overlapping
- Making it hard to separate the classes

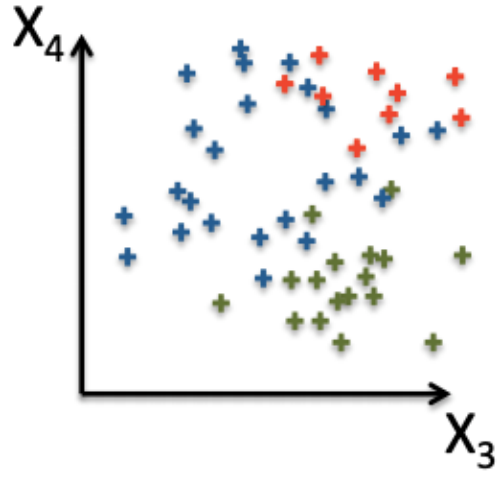


In Practice

- And we have many feature dimensions
- With some features more useful than others

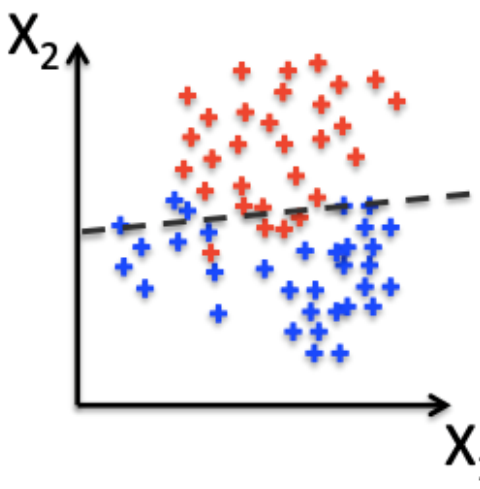






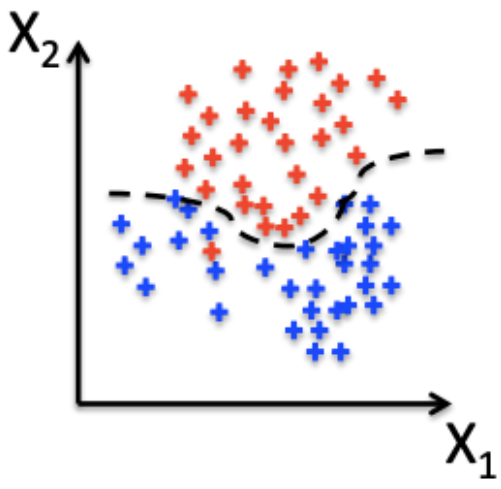
Different Models

- There are many different types of models that we can train to classify objects



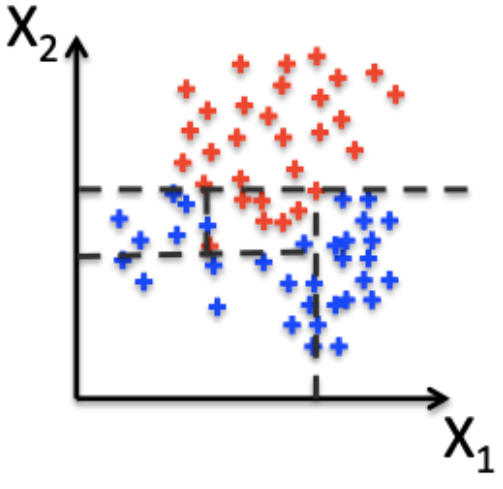
Linear classifiers

e.g. Logistic Regression, Linear SVMs



Non-linear Classifiers

e.g. Neural Nets, SVM with RBF kernel



Decision Tree Learners

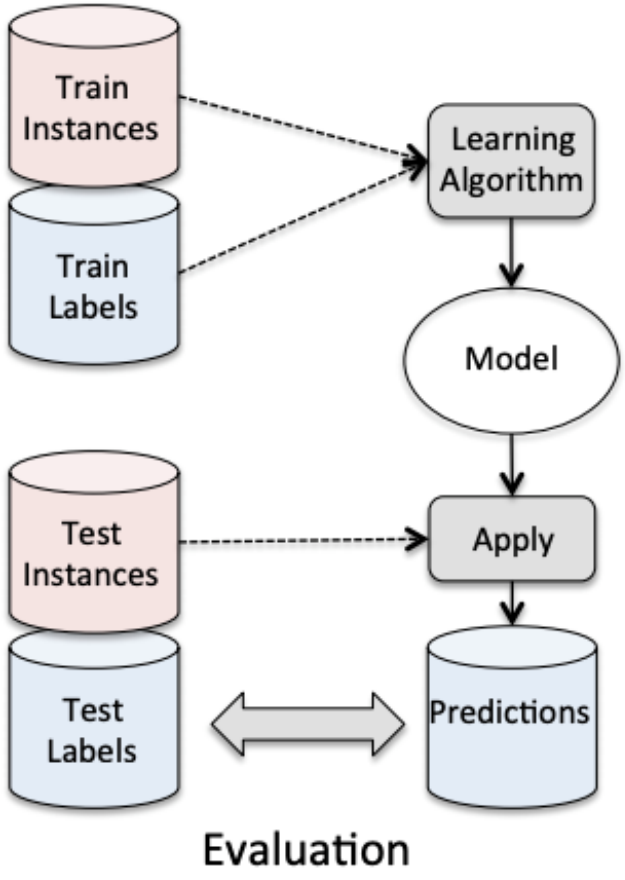
e.g. Random forests

FLUX Question

How can we decide which model is better?

Testing models

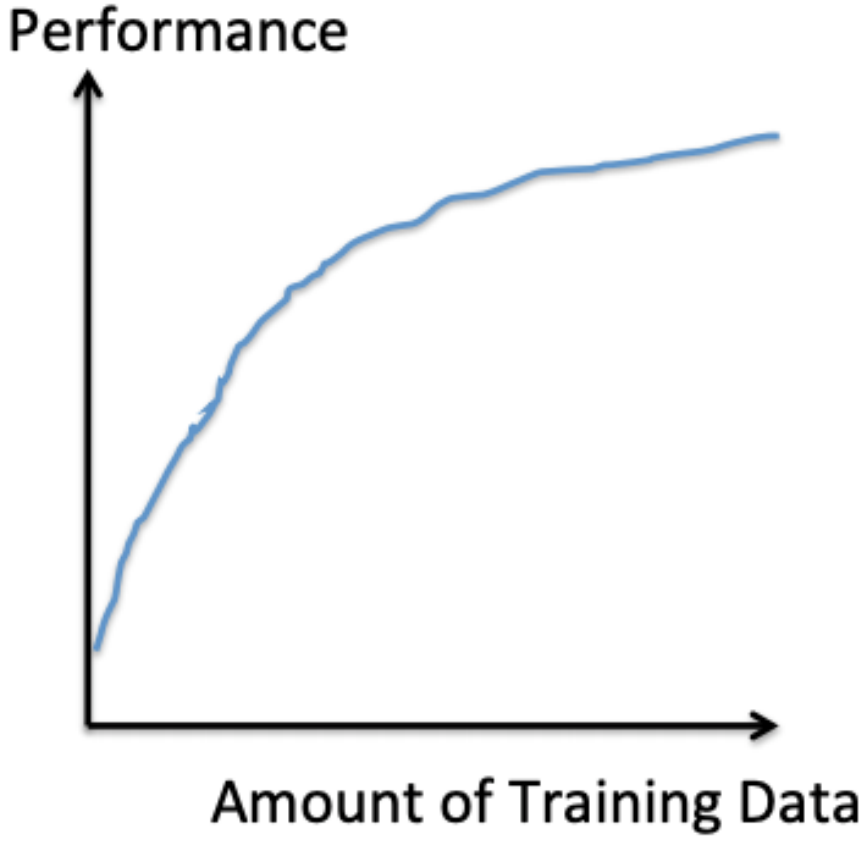
- We evaluate predictive models based on how well they predict the labels for test instances (not used in training)



Performance of predictive models

Generally:

- The more training data the better the test performance
- And (providing there is sufficient training data) the more features the better performance



End of Introduction

We'll talk more about predictive models in the coming weeks, especially in module 5.