INTRODUCTON TO EVERYTHING

Lecture 1

MAL1, 2025

this problem is solvable as there is an unambigious algorithm

$$\begin{array}{r} {\scriptstyle 1}\\ 129\\ +42\\ \hline 171 \end{array}$$

this problem is solvable as we've seen a lion before

CLASSIC ALGORITHM DESIGN

is based on unambiguous algorithms

CLASSIC ALGORITHM DESIGN

is based on unambiguous algorithms

MACHINE LEARNING

is based on learning from examples

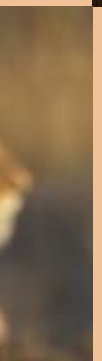






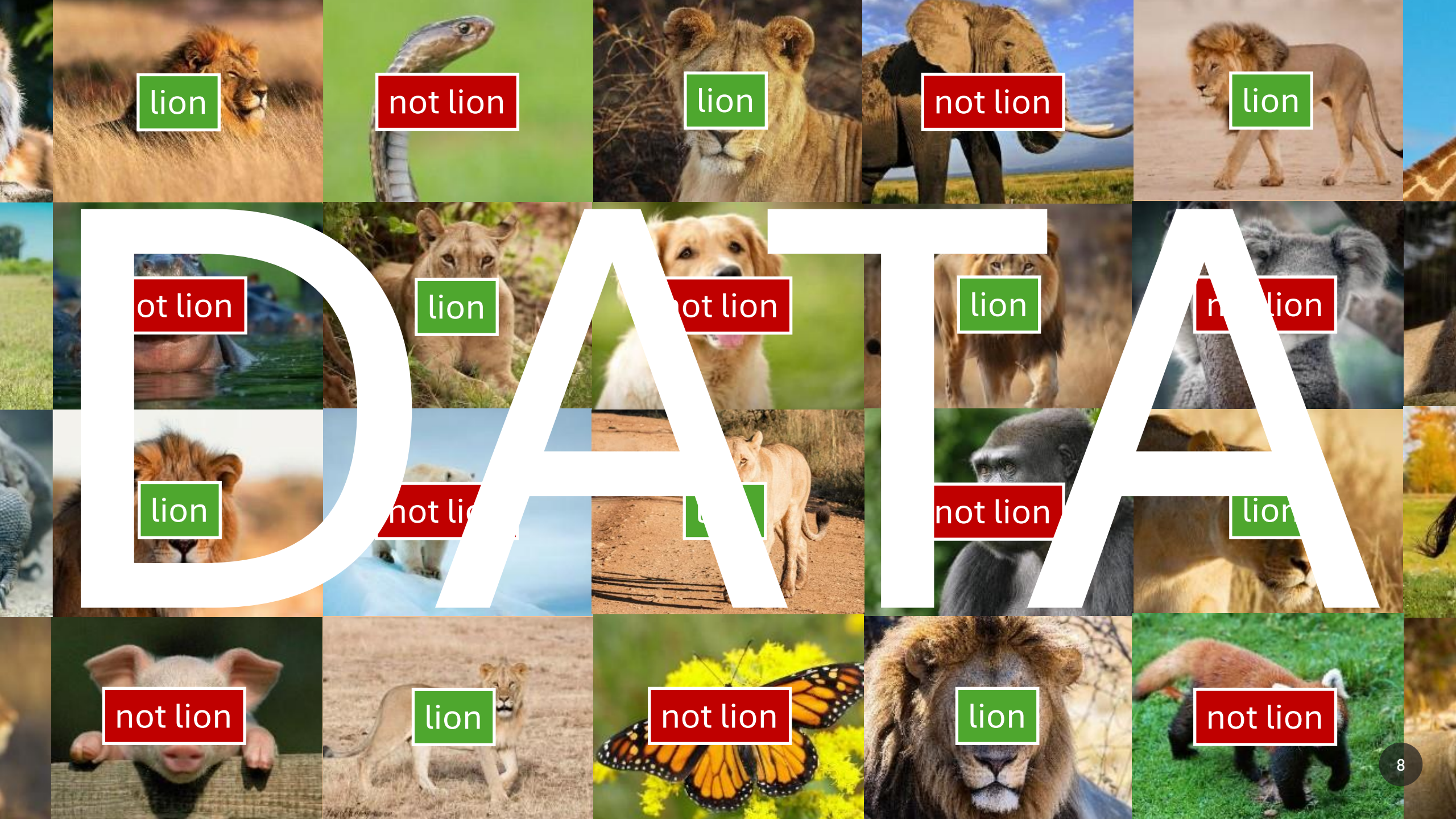


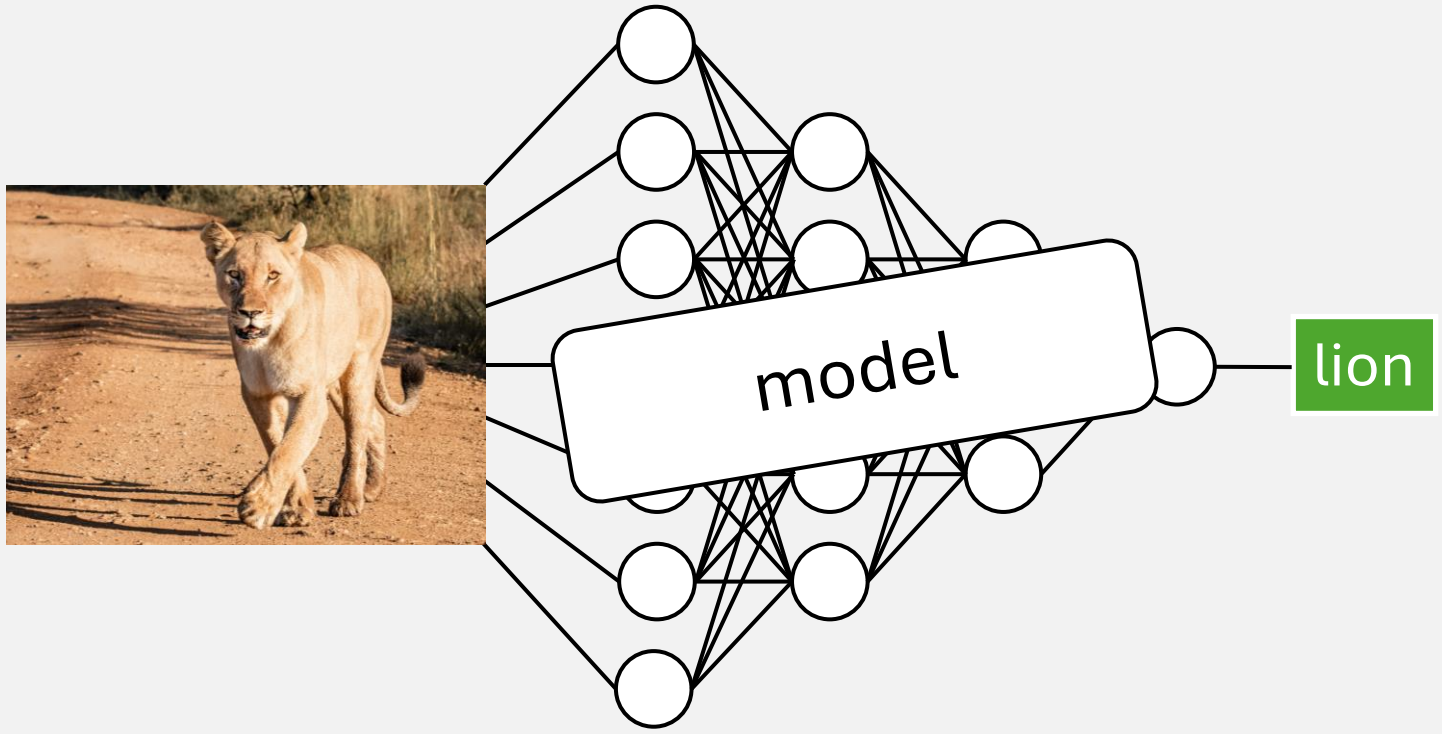




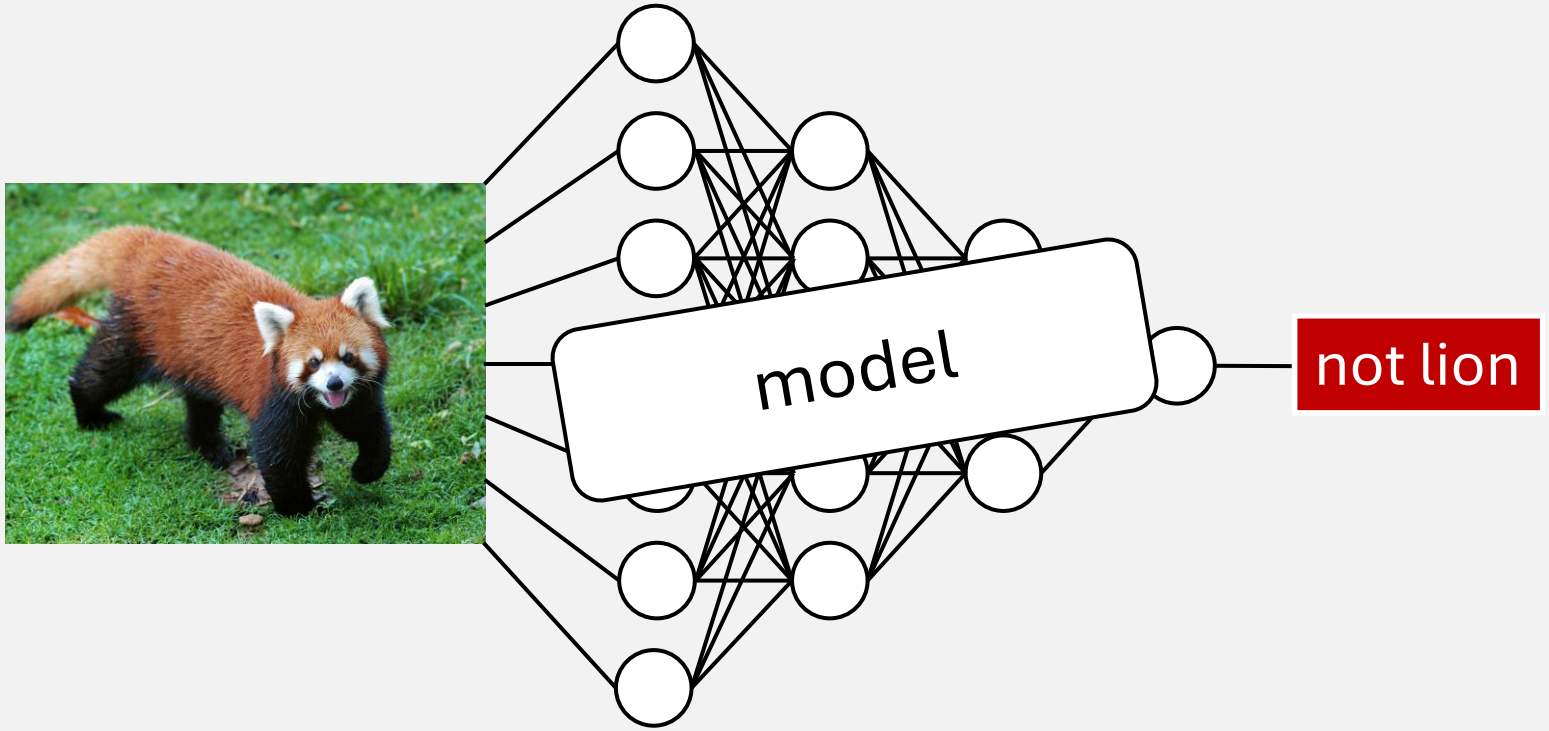


lion
not lion
lion
not lion
lion
not lion
not lion
lion
not lion
lion
not lion
not lion
lion
not lion
lion
not lion
lion
not lion
not lion
lion
not lion
lion
not lion

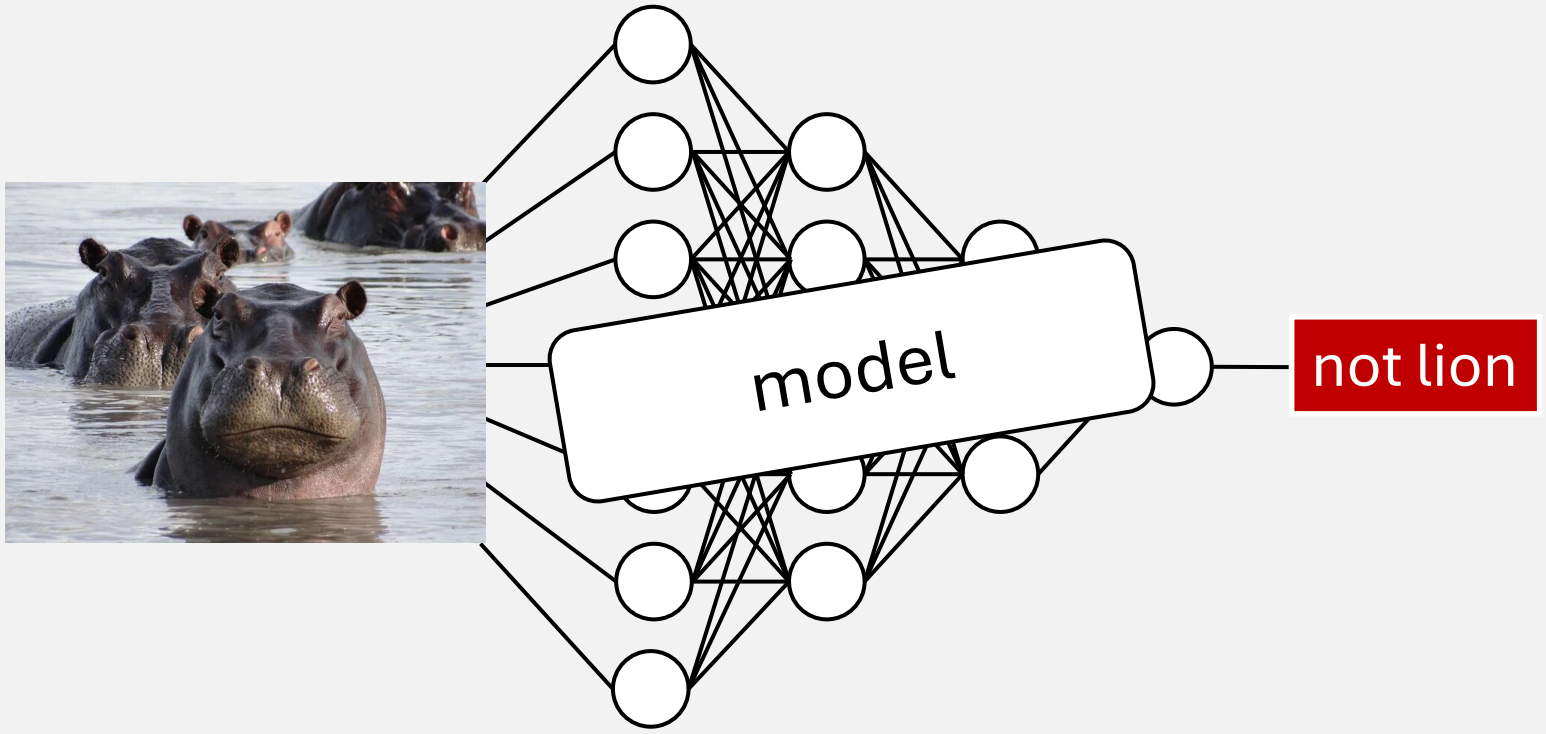

lion
not lion
lion
not lion
lion
ot lion
lion
ot lion
lion
lion
not lion
not lion
lion
not lion
lion
not lion
lion
not lion
DATA



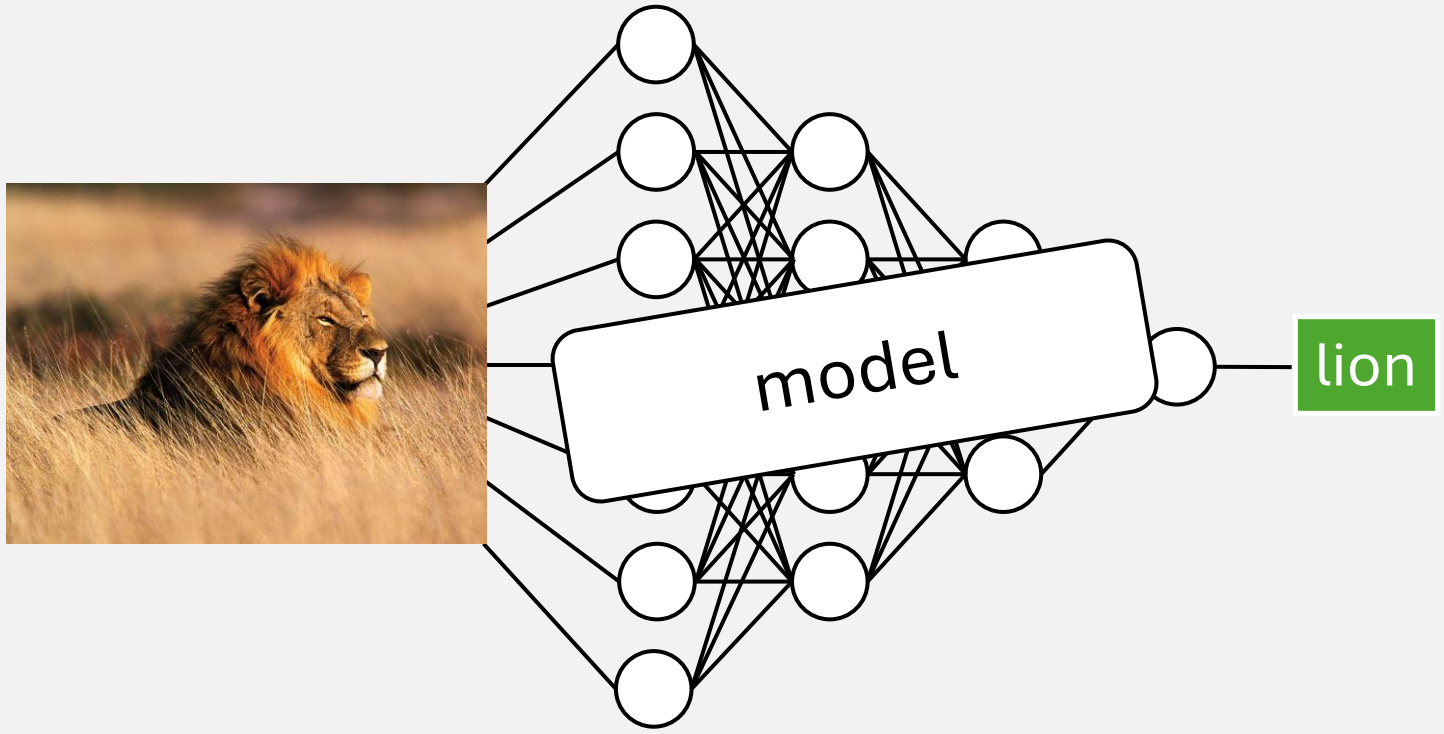
the model **learns from data**
and **writes the algorithm for us**



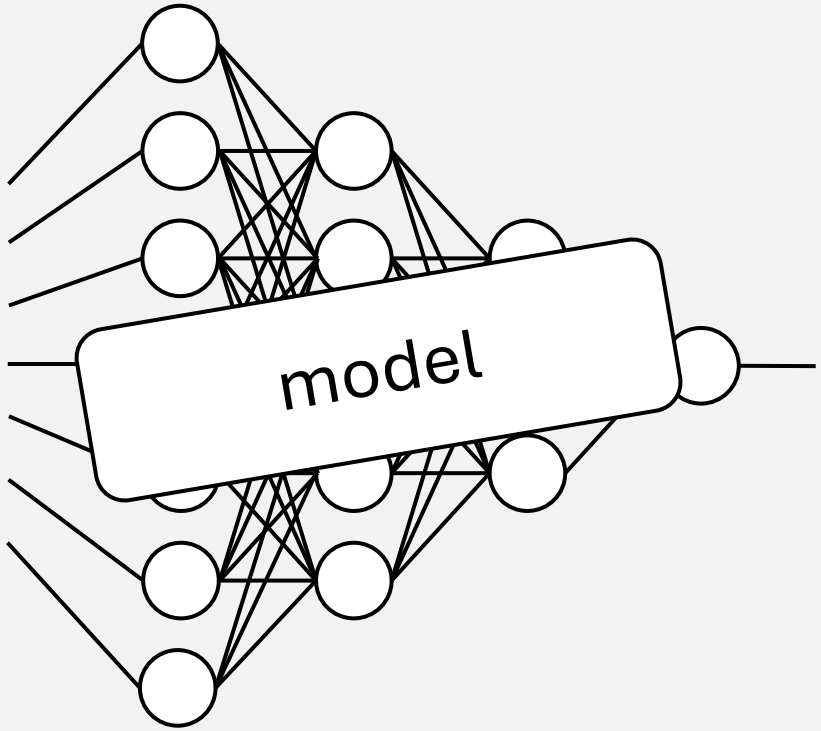
the model **learns from data**
and **writes the algorithm for us**



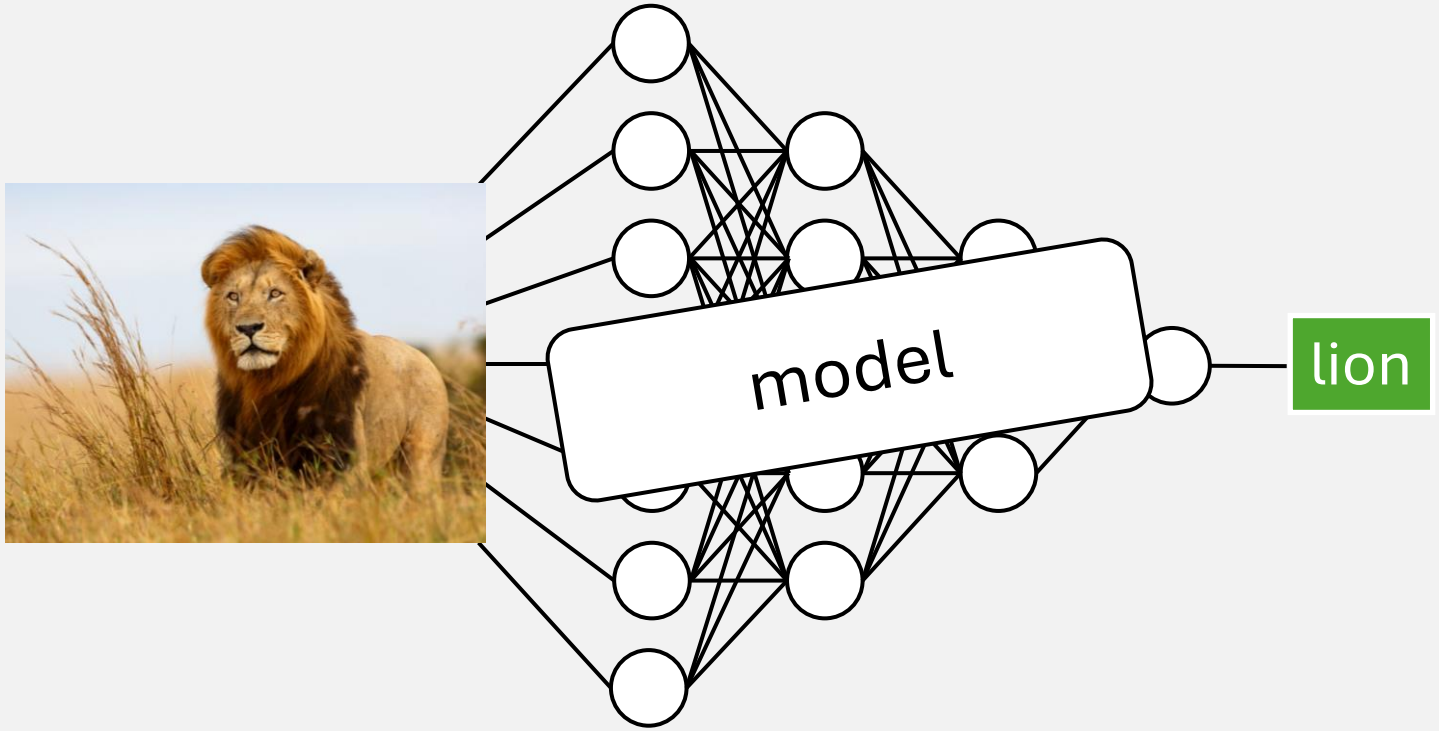
the model **learns from data**
and **writes the algorithm for us**

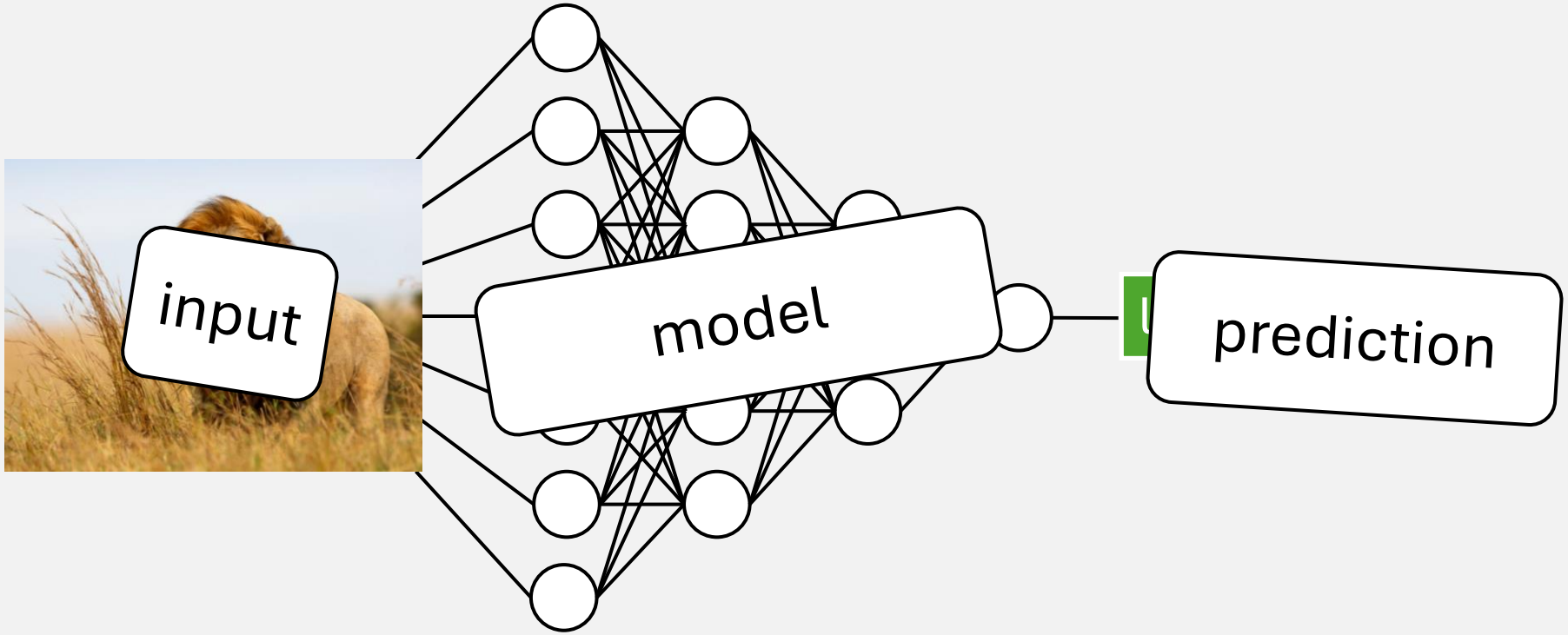


the model **learns from data**
and **writes the algorithm for us**

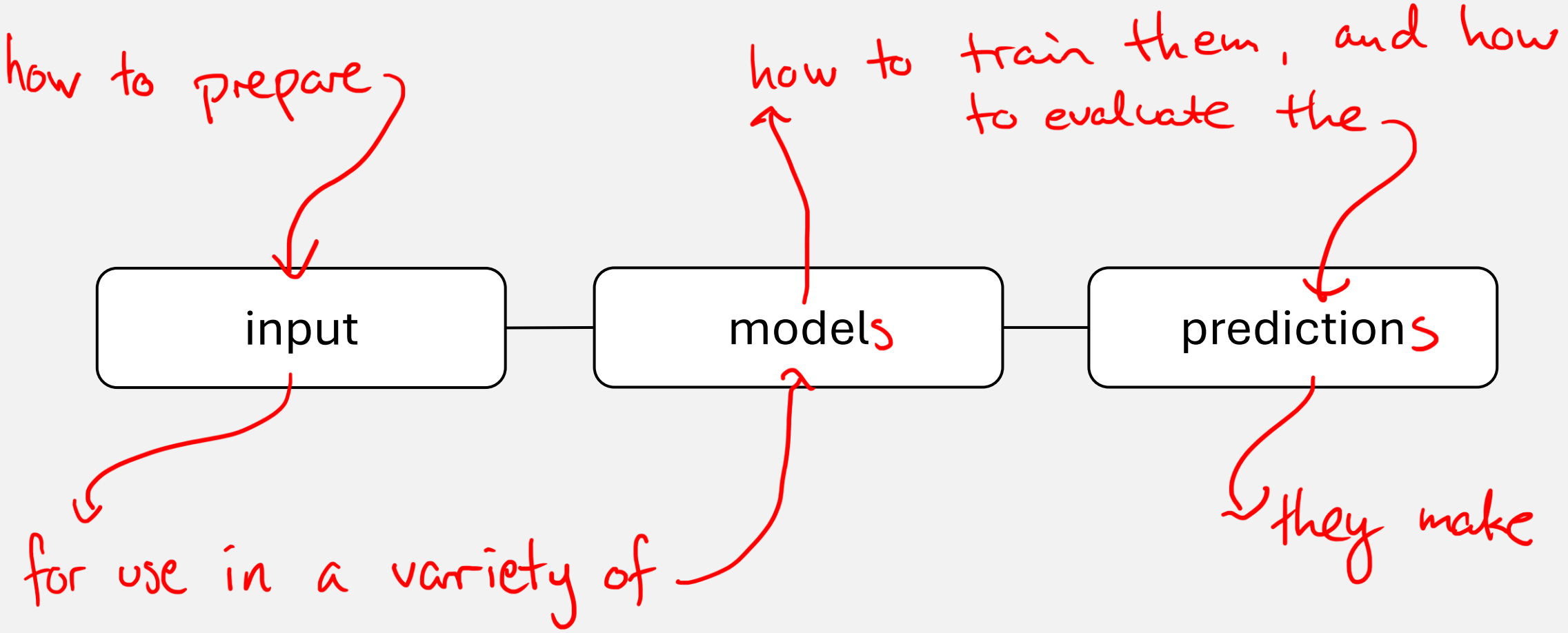


the model is **decoupled from data**
but the **algorithm is preserved**


model
lion


input
model
prediction

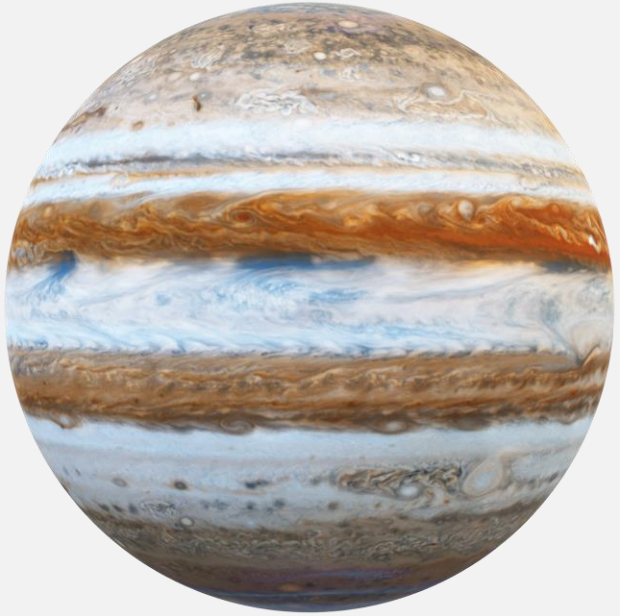
In this course, you will learn ...



INTRODUCTION TO EVERYTHING

- Who survived Titanic?
- The machine learning methodology
- Types of machine learning problems
- k-nearest neighbor
- Course overview

WHO SURVIVED TITANIC?

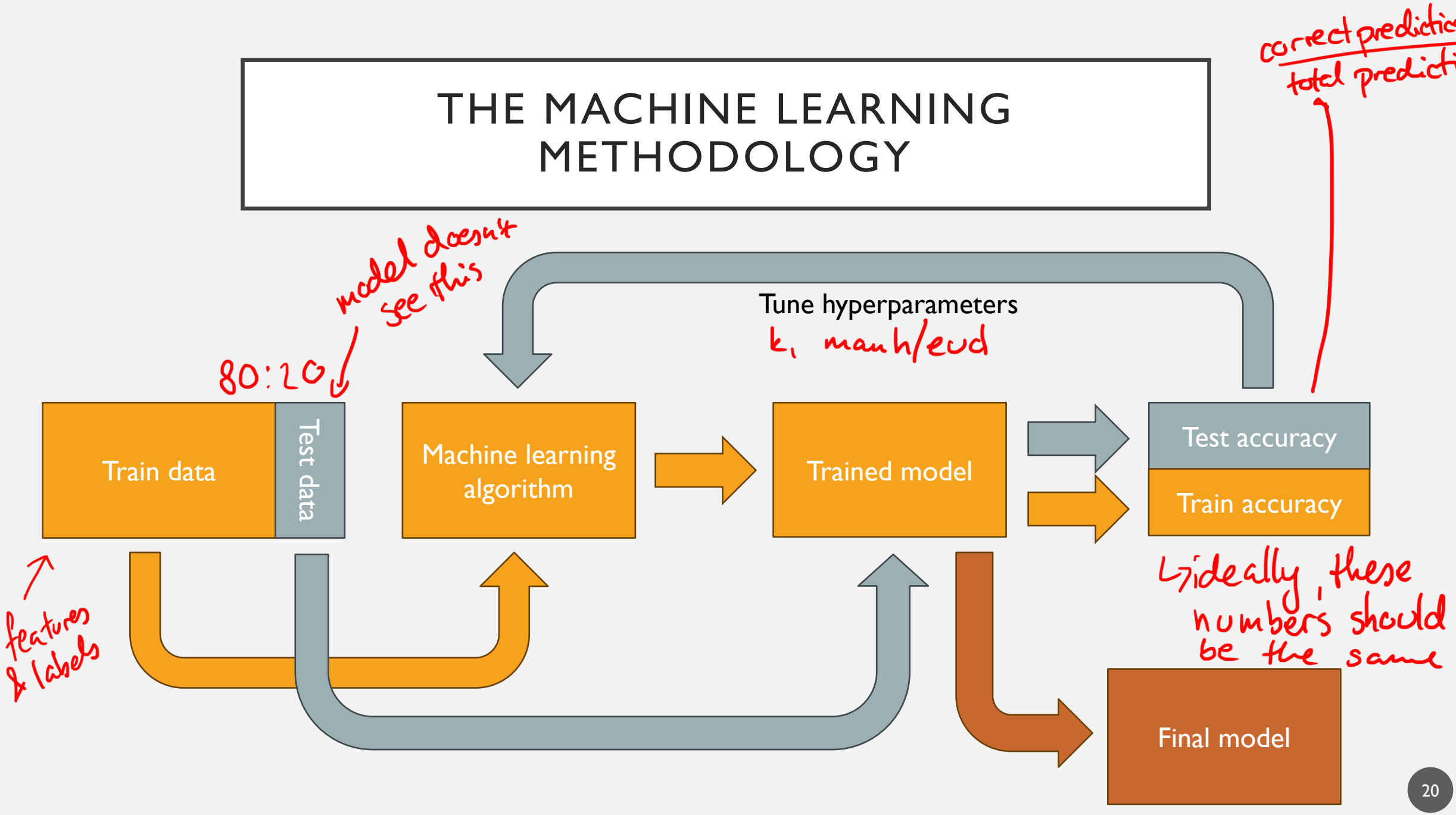


Jupyter Notebook *Titanic*

INTRODUCTION TO EVERYTHING

- Who survived Titanic?
- The machine learning methodology
- Types of machine learning problems
- k-nearest neighbor
- Course overview

THE MACHINE LEARNING METHODOLOGY



INTRODUCTION TO EVERYTHING

- Who survived Titanic?
- The machine learning methodology
- Types of machine learning problems
- k-nearest neighbor
- Course overview

THE DATA WE HAVE

LABELED

we have both  and lion

UNLABELED

we only have 

THE DATA WE HAVE

lion

Variables
Predictors
Covariates
Regressors
Dimensions
Attributes
Features
Inputs
Independent Variables

Response Variable
Outcome
Target
Class
Label
Output
Dependent Variable

TYPES OF MACHINE LEARNING PROBLEMS

LABELED DATA

SUPERVISED LEARNING

CLASSIFICATION

y is a category

REGRESSION

y is a number

predictions

UNLABELED DATA

UNSUPERVISED LEARNING

CLUSTERING

identify groups in data

DIMENSIONALITY REDUCTION

identify structure in data

patterns

CLASSIFICATION

fraud detection
speech recognition
diagnostics

REGRESSION

weather forecasts
stock market predictions
dependency insights

CLUSTERING

targeted marketing
anomaly detection
image segmentation

DIMENSIONALITY REDUCTION

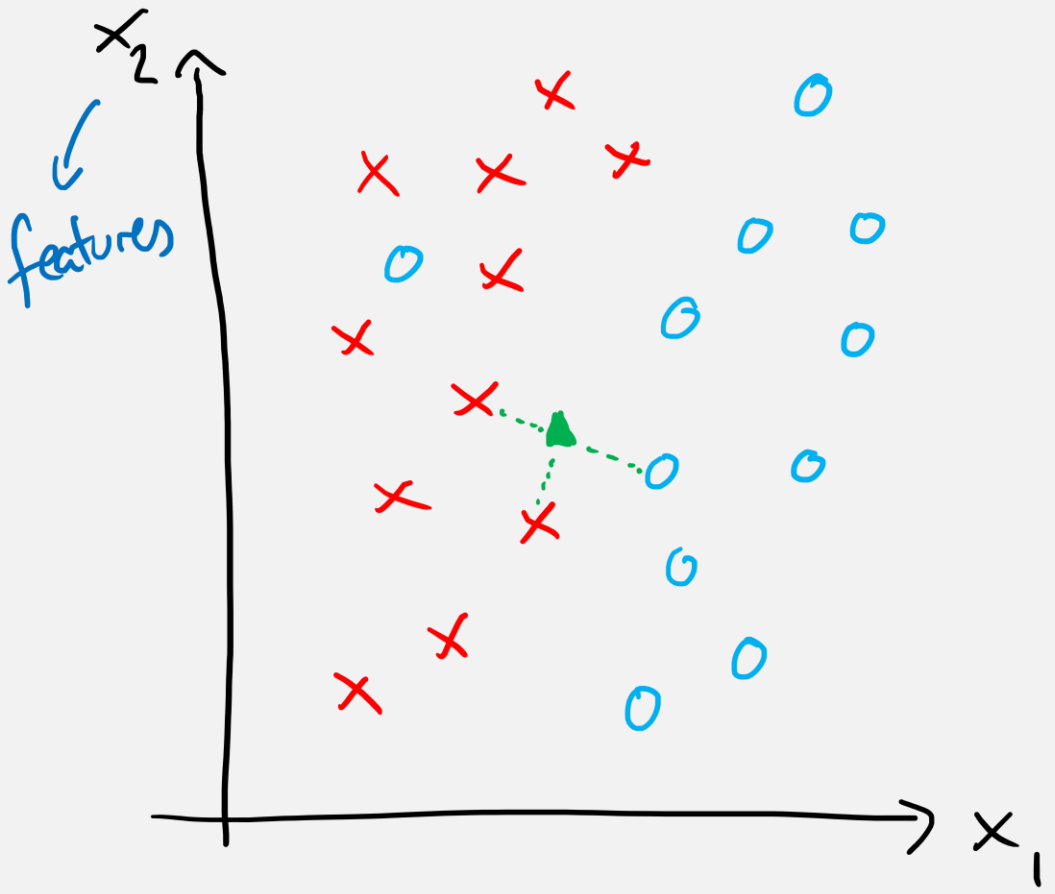
compression
structure discovery
big data visualization

INTRODUCTION TO EVERYTHING

- Who survived Titanic?
- The machine learning methodology
- Types of machine learning problems
- k-nearest neighbor
- Course overview

k-NEAREST NEIGHBOR

a classifier

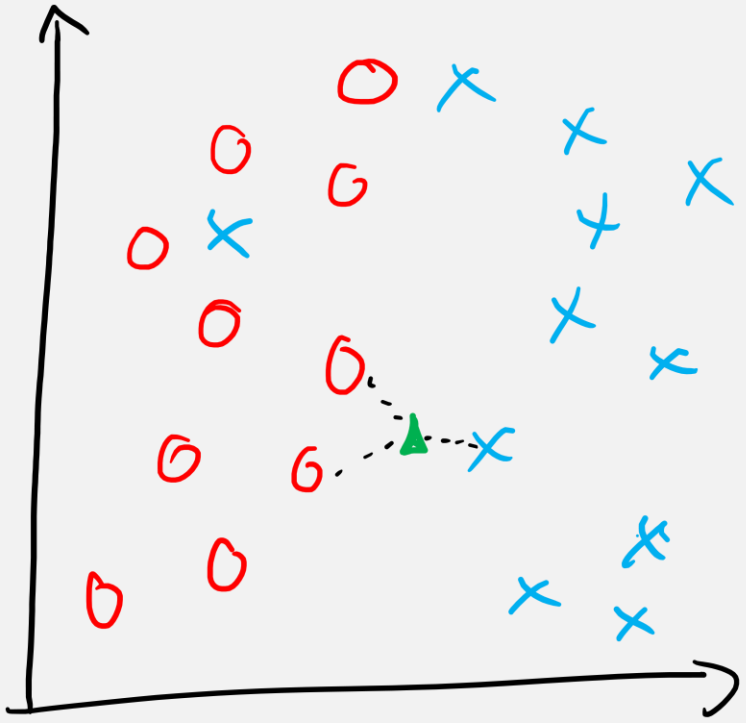


find the k (=3) nearest neighbours and assign to majority class

Hyperparameters

1. The value of k
2. How distance is measured

THE VALUE OF *k*



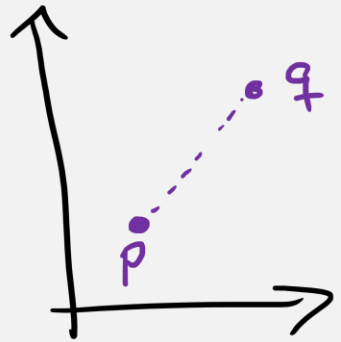
$k=1$ ▲ = X

$k=3$ ▲ = O

THE DISTANCE METRIC

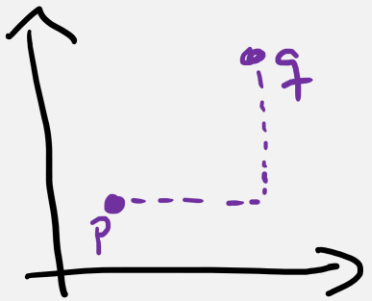
Euclidean

$$d(p,q) = \sqrt{(p_1-q_1)^2 + (p_2-q_2)^2 + \dots}$$



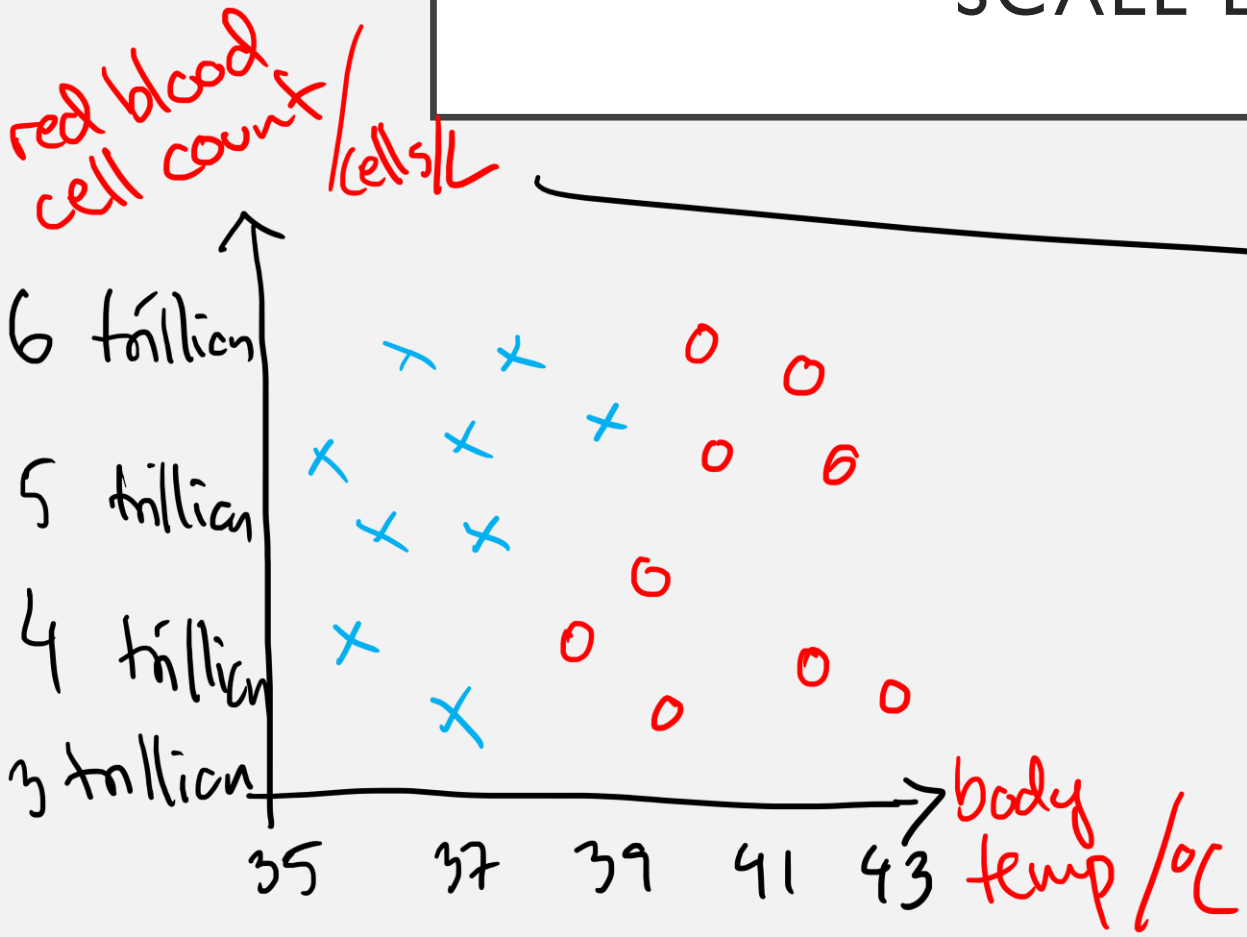
Manhattan

$$d(p,q) = |p_1-q_1| + |p_2-q_2| + \dots$$



+ Minkowski, Hamming, Levenshtein, Mahalanobis …

SCALE EFFECTS



this will have a much larger influence on distance between samples, biasing the classifier

You will learn much more about both how to tune hyperparameters and how to deal with scale effects later in the course

INTRODUCTION TO EVERYTHING

- Who survived Titanic?
- The machine learning methodology
- Types of machine learning problems
- k-nearest neighbor
- Course overview

COURSE OVERVIEW

What?

12 sessions

6 portfolio assignments

1 oral exam

the heart and soul of the course

Who?

Frederik Thorning Bjørn
frbj@via.dk
Lessons 1 and 3-9

will also be your MAL1-in-SEP4 supervisor

Richard Brooks
rib@via.dk
Lessons 2 and 10-12

COURSE OVERVIEW

1. Introduction to everything
2. Mathematical background
3. Regression
4. Preprocessing and feature engineering
5. Tree-based methods
6. Validation methods and performance metrics
7. Logistic regression and gradient descent
8. Support vector machines
9. Neural networks
10. Dimensionality reduction
11. Clustering
12. Recap

Portfolio assignment n
will be handed out in session $2n - 1$
and handed in before session $2n + 1$

- Explain what machine learning is
- Distinguish between supervised and unsupervised learning
- Explain and implement the kNN classifier