# Machine Learning

## Lecture 1 – Introduction to Machine Learning

**Dr SHI Lei**

# Today

- Machine Learning

  What is this module about?

- General Module Information

# What is Machine Learning?

# What is Machine Learning?

ML is a “field of study that gives computers the ability to learn without being explicitly programmed”.

-- Arthur Samuel, 1959

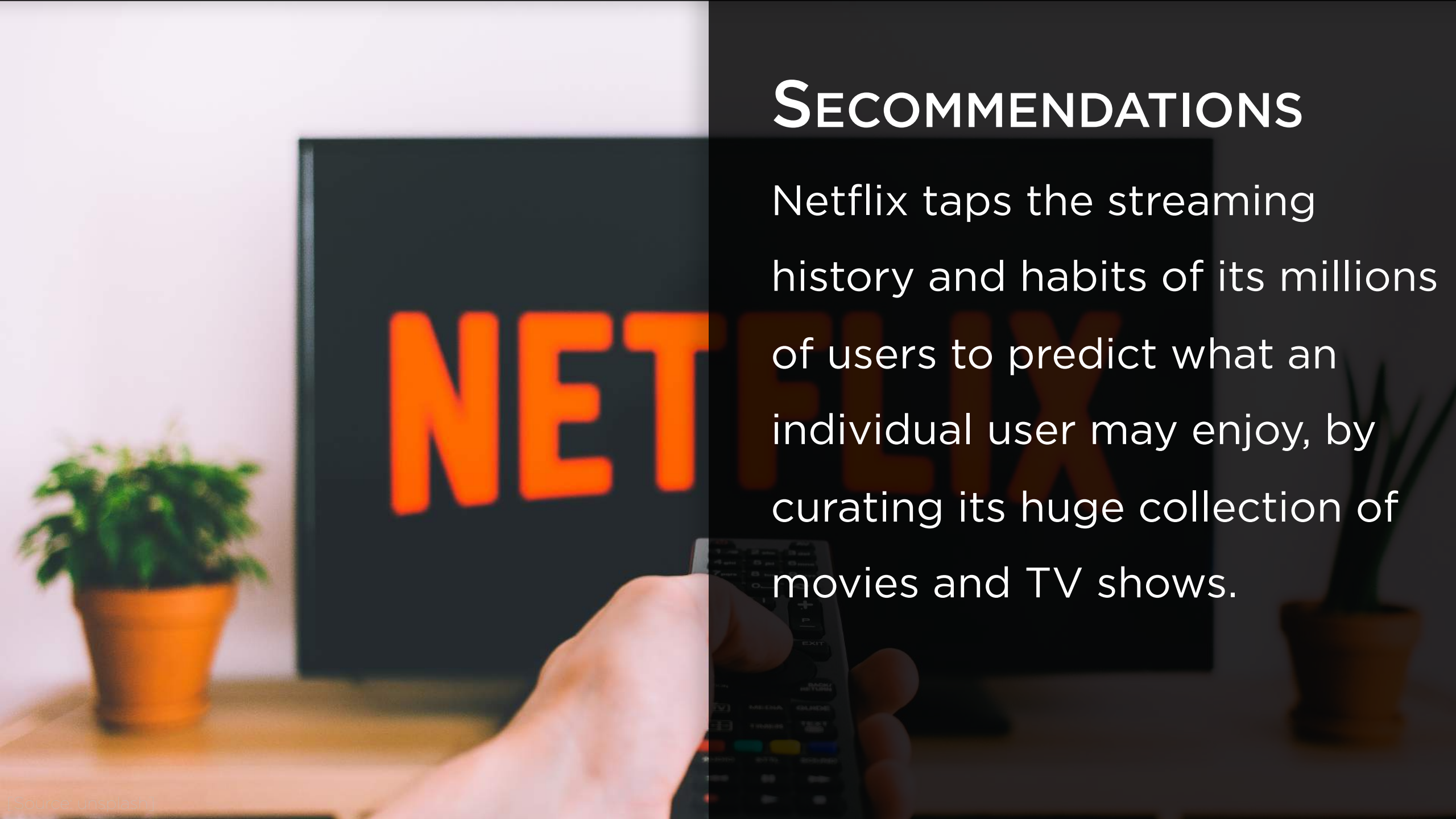
# What is Machine Learning?

ML will allow you, as a software engineer, to do 3 things:

1. to reduce the time you spend programming;
2. to customise product, making it better for specific users;
3. to solve problems that you, as a programmer, have no idea how to do by hand.

# What is Machine Learning?

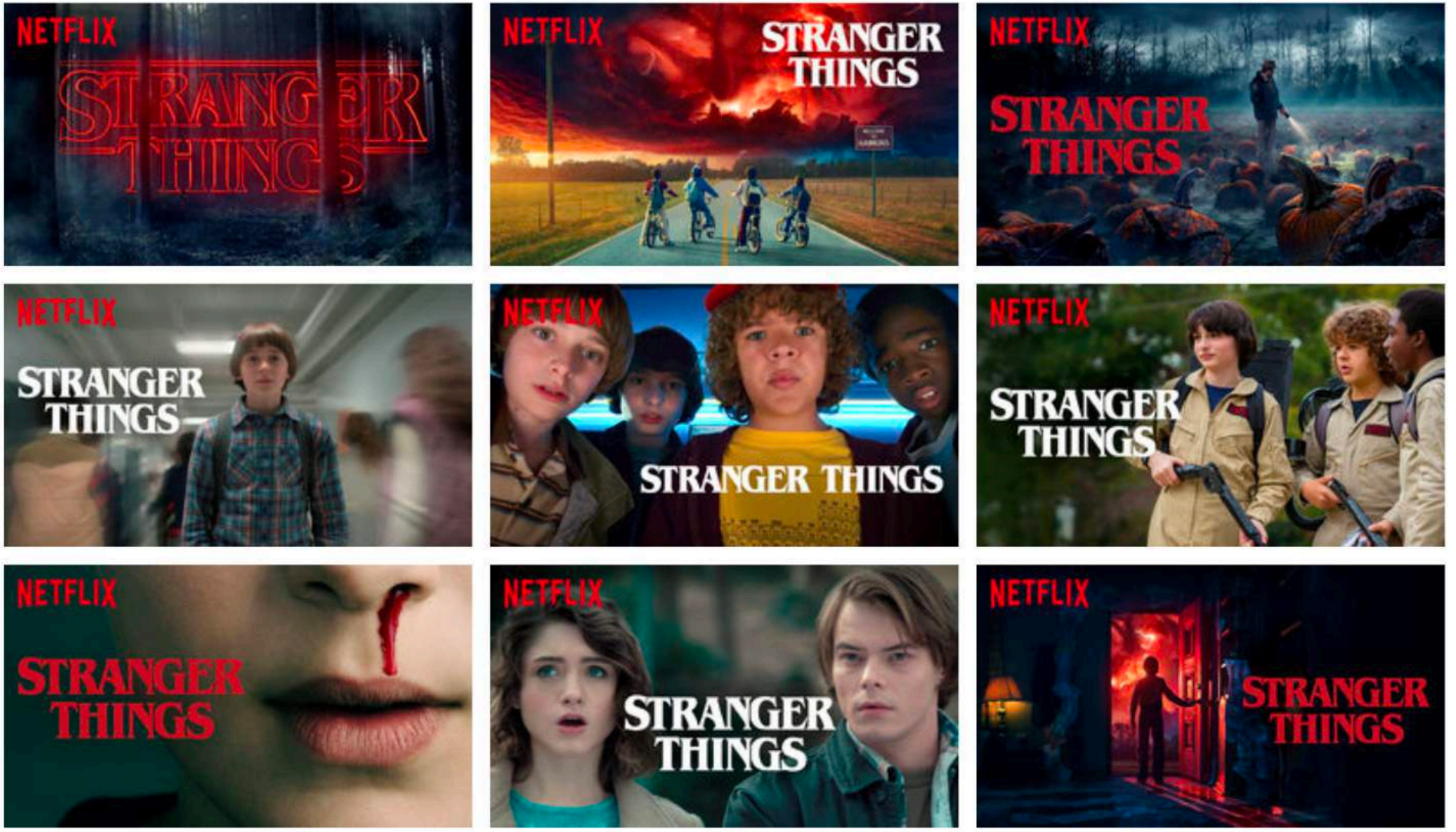
Some ML examples...




# SECOMMENDATIONS

Netflix taps the streaming history and habits of its millions of users to predict what an individual user may enjoy, by curating its huge collection of movies and TV shows.

Find the single perfect artwork for **Stranger Things** across all Netflix members.




[Source: Netflix Techblog]


# Relevant Timelines

Twitter predicts how interesting and engaging a Tweet would be specifically to an individual user, in order to rank Tweets based on relevance thus provide personalised timelines.

[Source: freestocks]


[Source: REUTERS/Stringer]

# Optimised Logistic

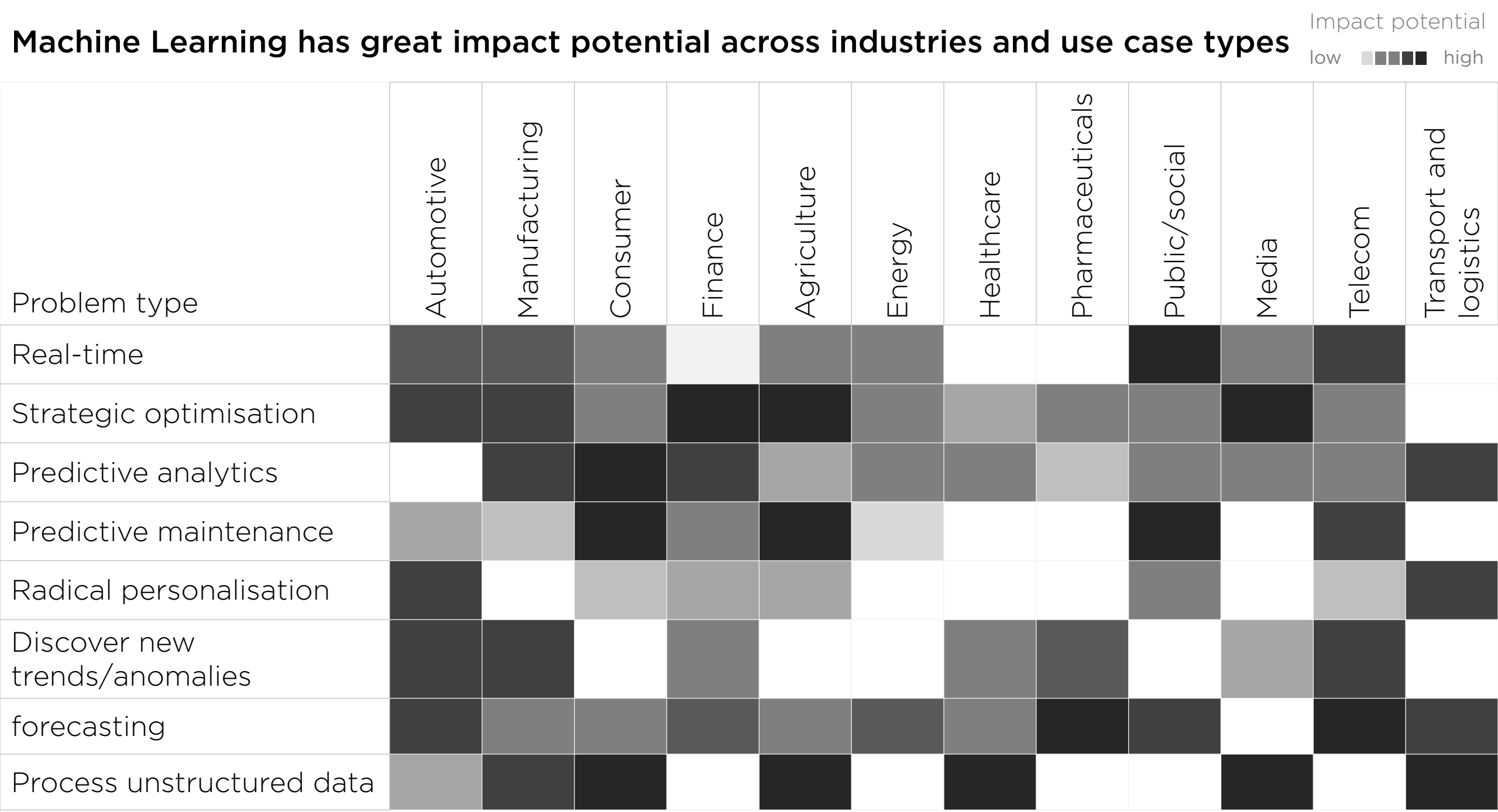
Alibaba uses purchase histories to generates precise order predictions which help online sellers to pre-load their warehouses with certain amounts of inventories.


# Self-driving Cars

Waymo's self-driving cars scan constantly for objects around, and predicts their possible paths to determine the exact trajectory, speed, lane, and steering manoeuvres needed to progress along the route safely.

[Source: Waymo]

# Machine Learning has great impact potential across industries and use case types



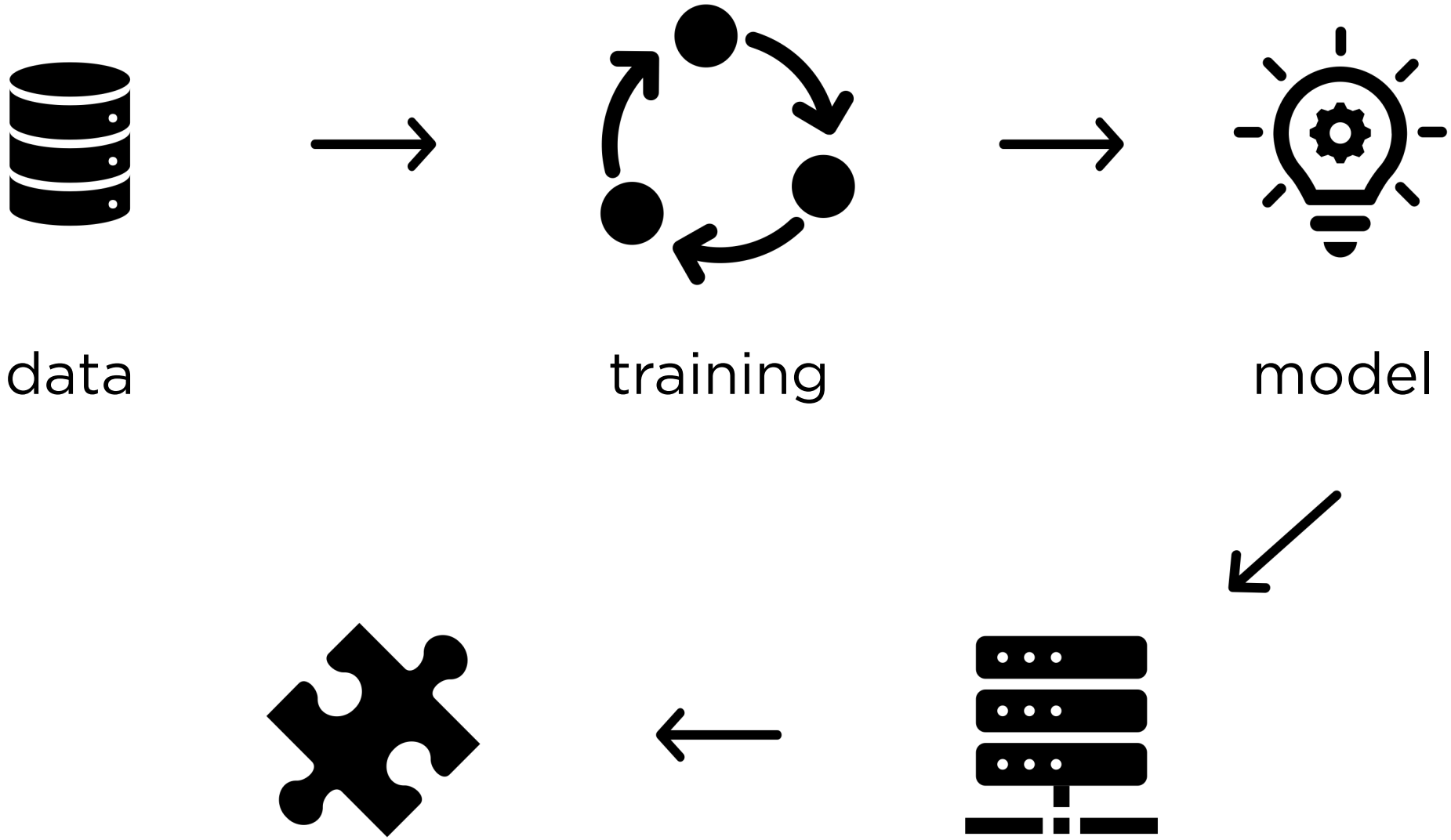

[Source: Théo Szymkowiak]

# 7 steps in Machine Learning Lifecycle


1. Gathering data
2. Preparing data
3. Choosing model
4. Training
5. Evaluation
6. Hyper-parameter tuning
7. prediction

Training Prediction

Using data to answer questions


data
training
model
serve predictions
answer questions

So, how does Machine Learning really work?

# Green tea / Oolong tea

[Source: oriento]

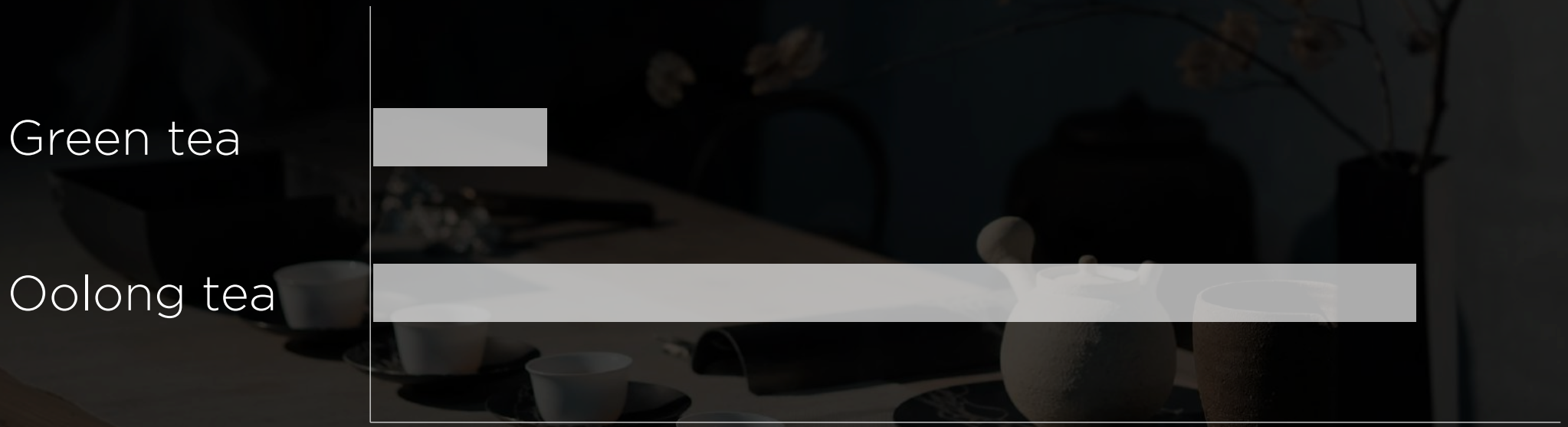
# 1. Gathering the Data

| Caffeine (mg) | Acidity (PH level) | Green tea or oolong tea? |
|---|---|---|
| | | |
| | | |
| | | |
| | | |
| | | |

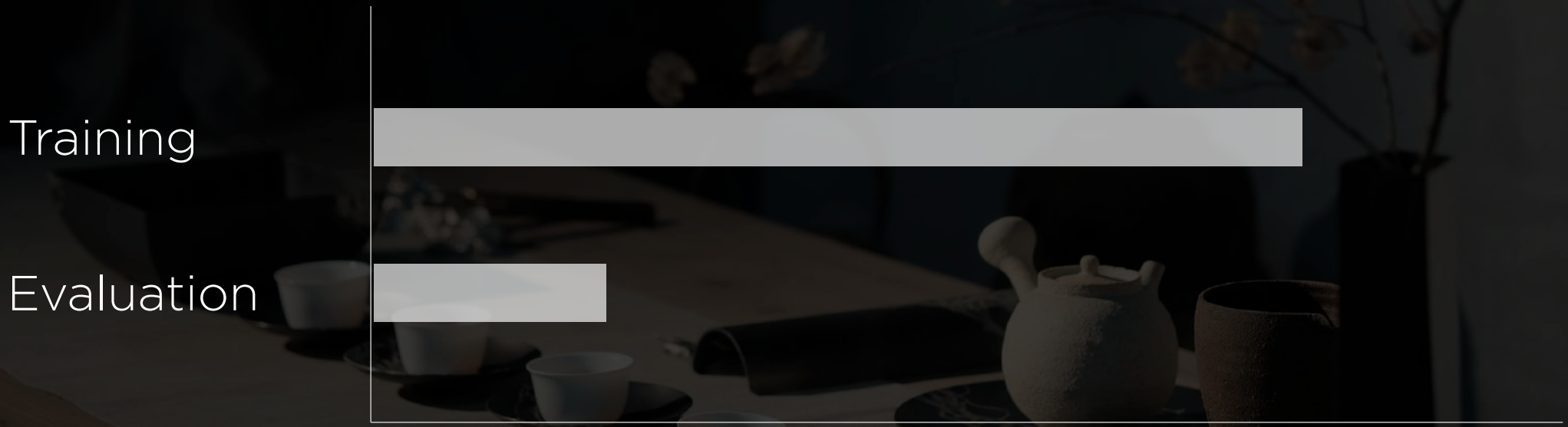
# 2. Data Preparation

| Caffeine (mg) | Acidity (PH level) | Green tea or oolong tea? |
|---|---|---|
| 50 | 5.8 | Green tea |
| 65 | 6.2 | Oolong tea |
| 40 | 7.9 | Green tea |
| 60 | 6.5 | Green tea |
| 70 | 7.8 | Oolong tea |

# 2. Data Preparation



# 2. Data Preparation



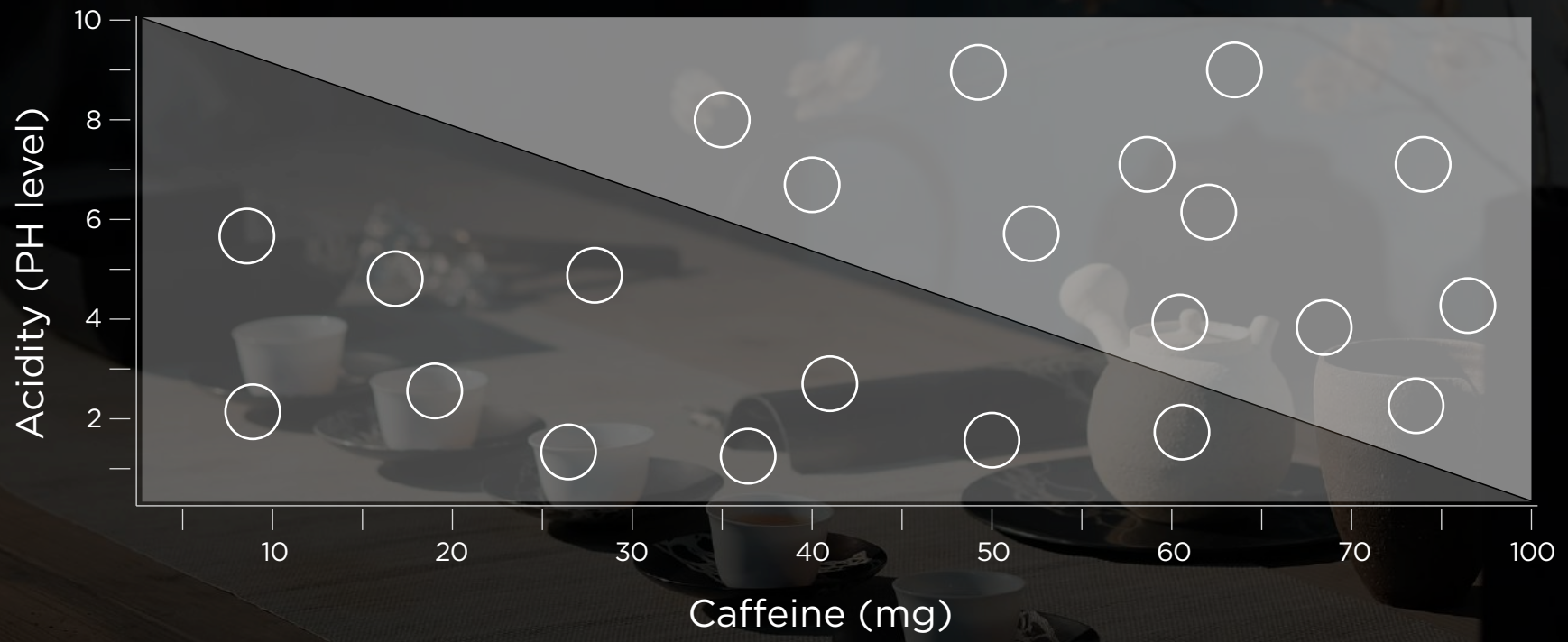
Green tea / Oolong tea

# 3. Choosing a Model



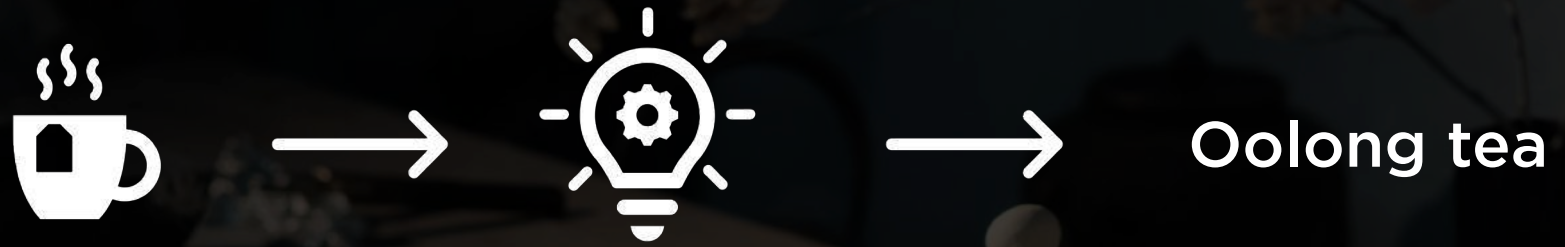

Green tea / Oolong tea

# 3. Choosing a Model



Green tea / Oolong tea

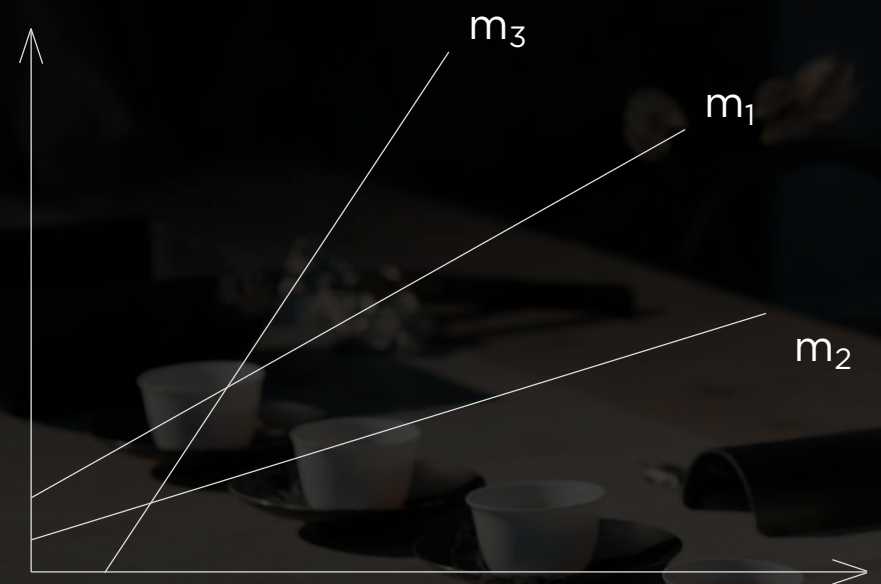
# 4. Training



model

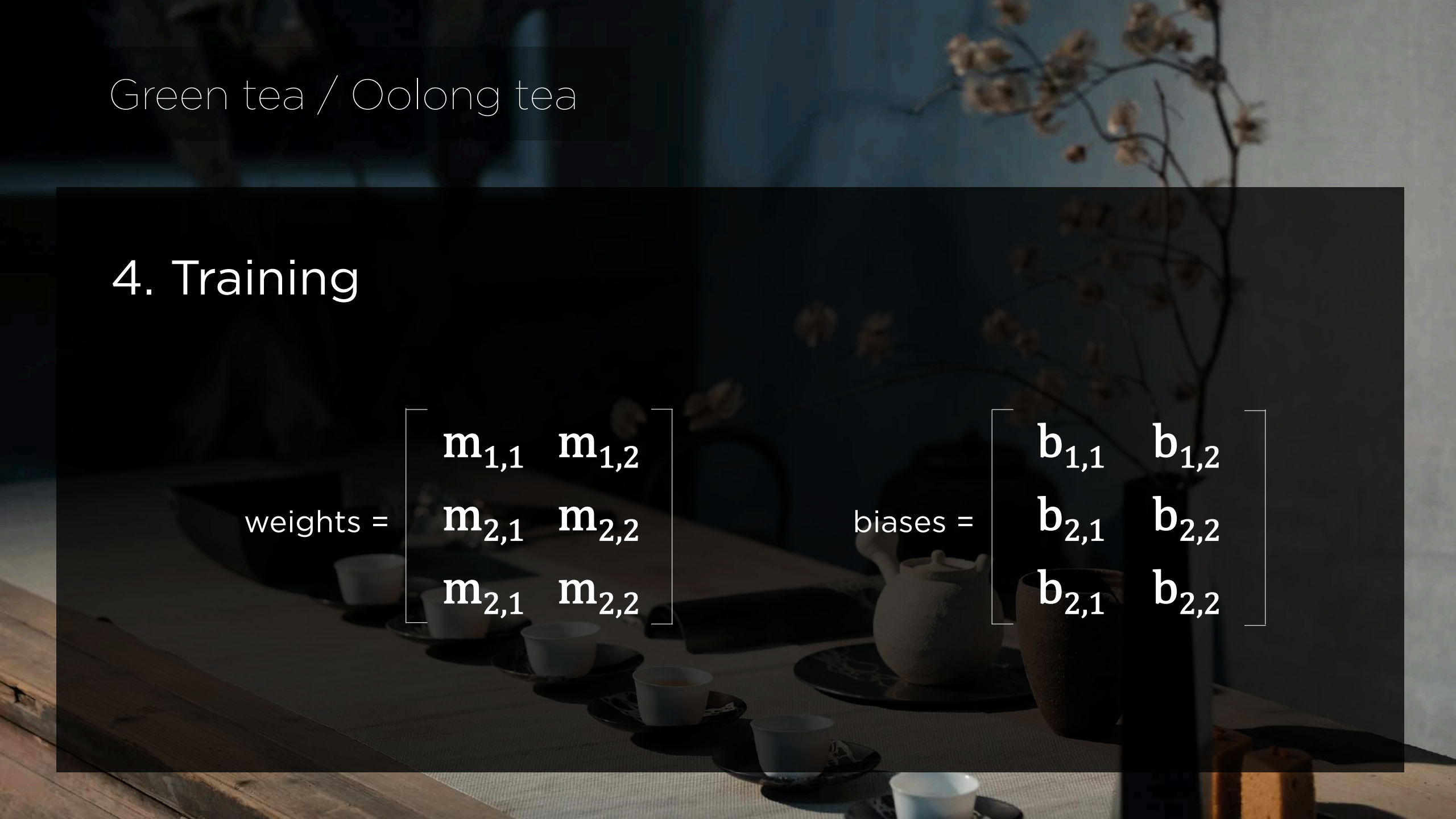
Oolong tea

# 4. Training




$$y = m \times x + b$$

output slope input y-intercept

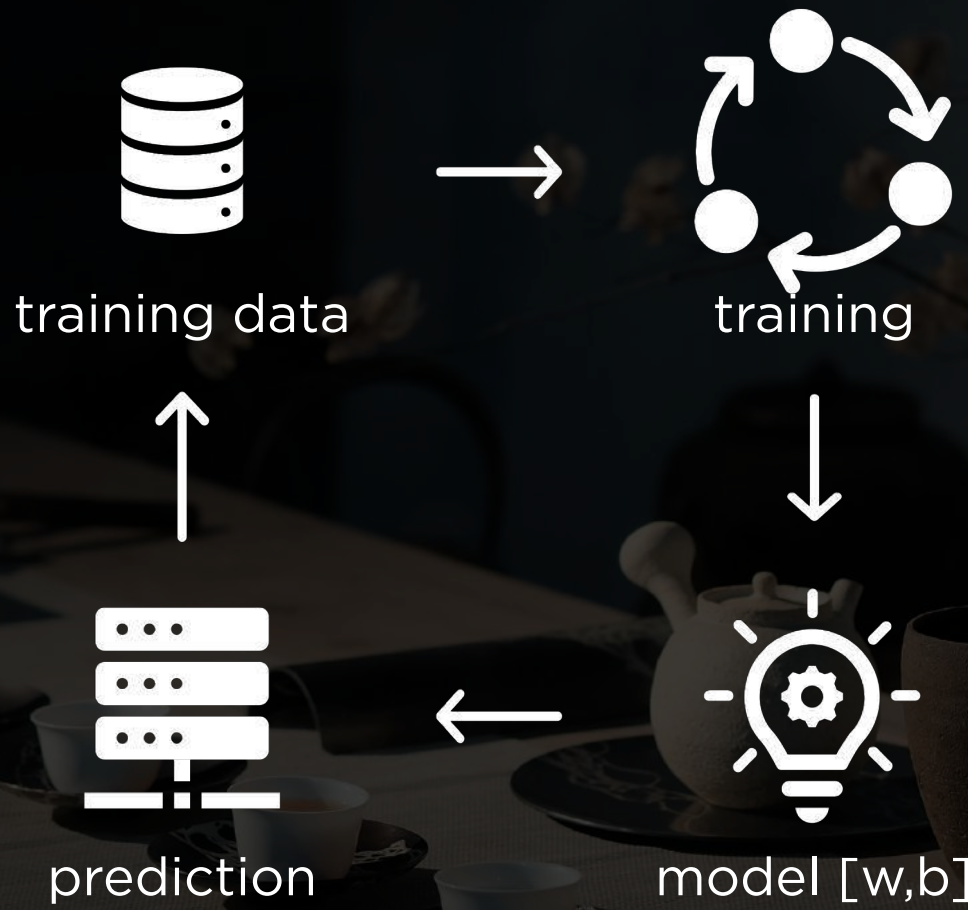



# 4. Training

$$\text{weights} = \begin{bmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \\ m_{2,1} & m_{2,2} \end{bmatrix} \qquad \text{biases} = \begin{bmatrix} b_{1,1} & b_{1,2} \\ b_{2,1} & b_{2,2} \\ b_{2,1} & b_{2,2} \end{bmatrix}$$

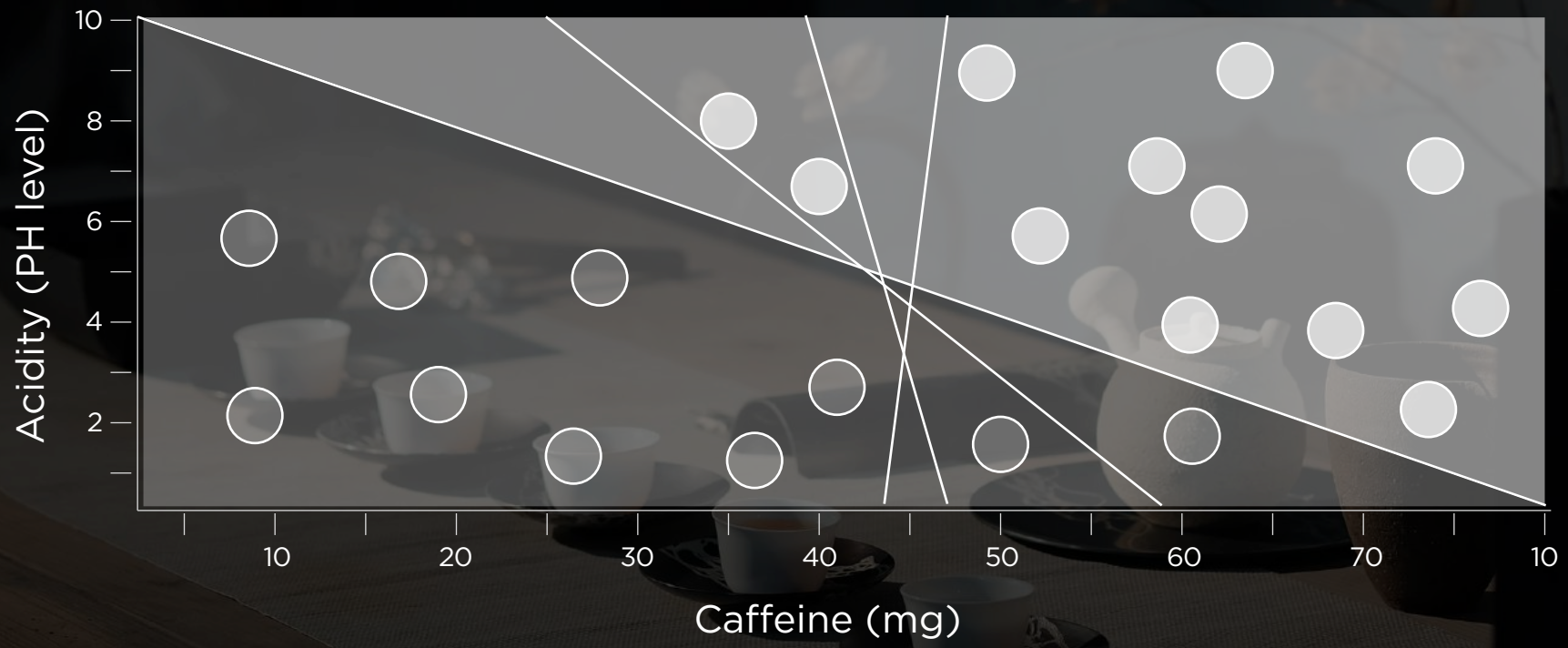
Green tea / Oolong tea

# 4. Training



Green tea / Oolong tea

# 4. Training

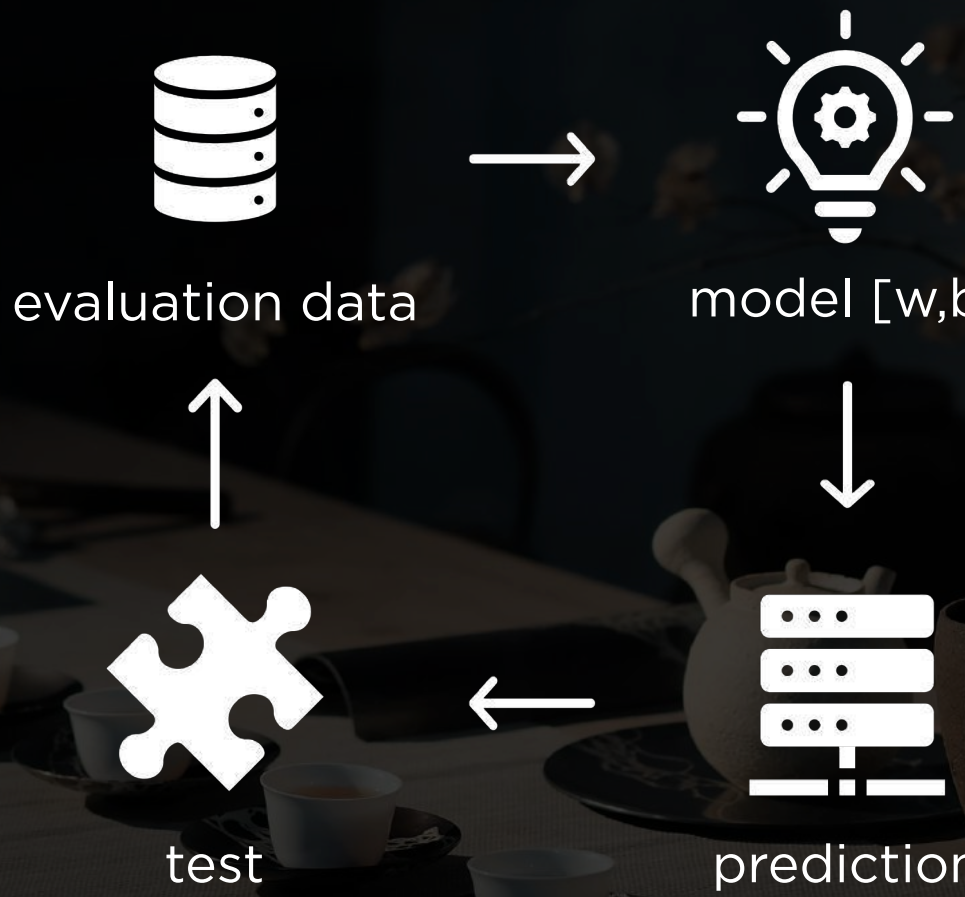


# 5. Evaluation

evaluation data

Green tea / Oolong tea

# 5. Evaluation



Green tea / Oolong tea

# 5. Evaluation

training data

80% / 70%

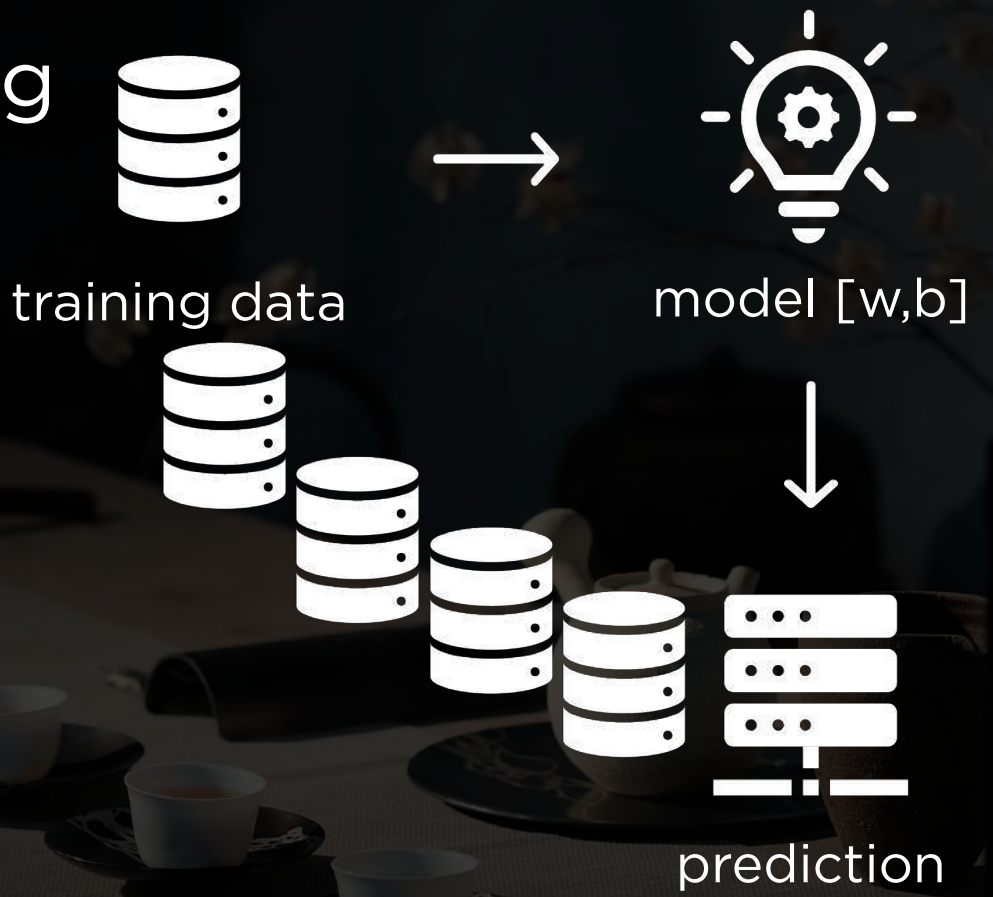
evaluation data

20% / 30%

The training-evaluation split depends on the size of dataset

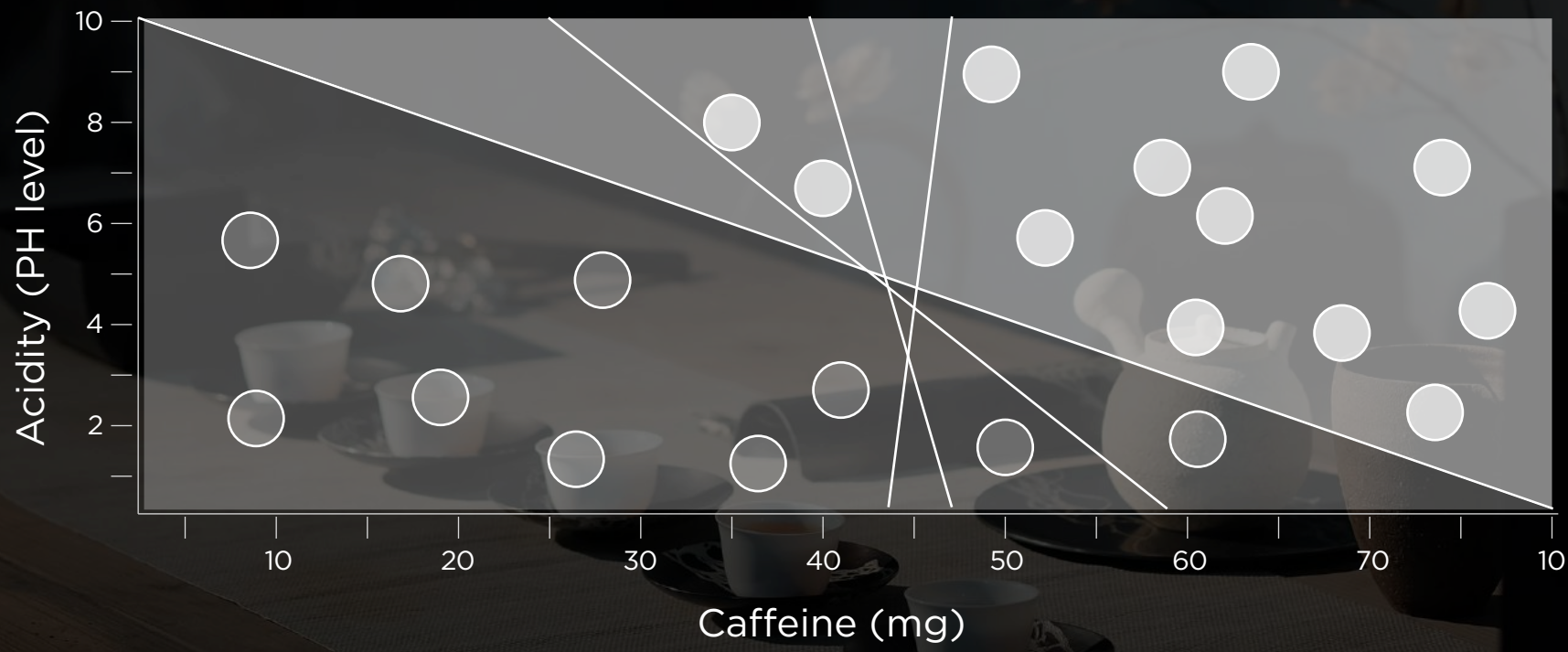
Green tea / Oolong tea

# 6. Parameter Tuning



Green tea / Oolong tea

# 6. Parameter Tuning

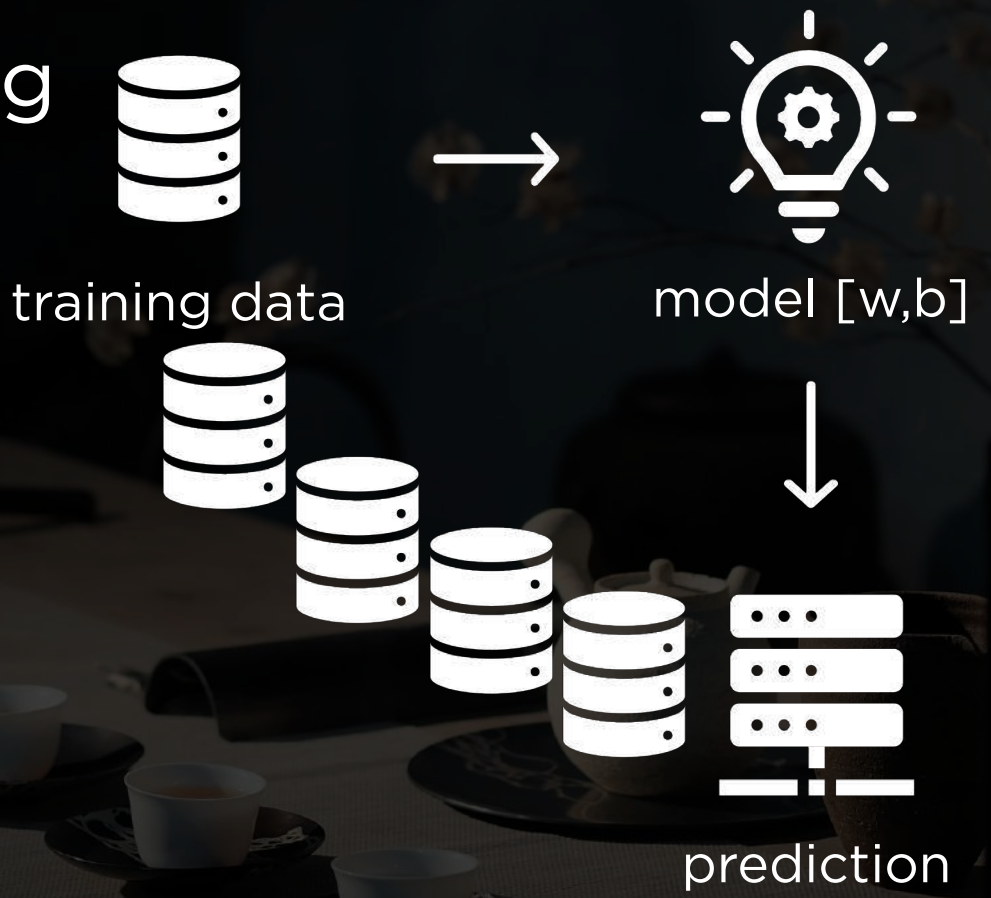



Learning rate

Green tea / Oolong tea

# 6. Parameter Tuning

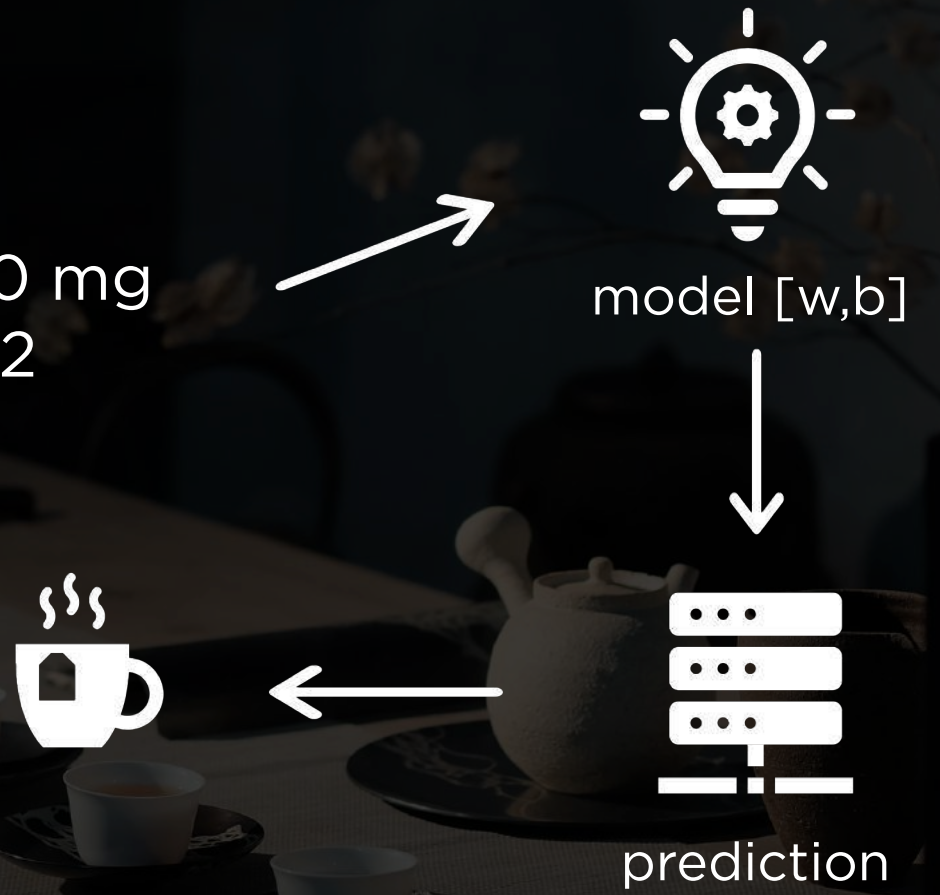
(hyperparameters)



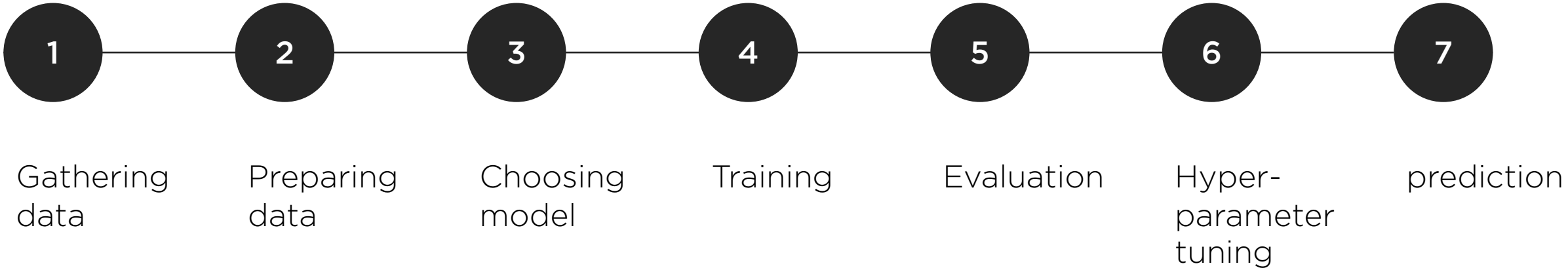
Green tea / Oolong tea

# 7. Prediction

Caffeine: 60 mg
Acidity: 7.2



# 7 steps in Machine Learning Lifecycle




- to determine how to differentiate between oolong tea and green tea using our model rather than using human judgement and manual rules.
- We can extrapolate the ideas to other problem domain as well, where the same principles apply – 7 steps...

# What is Machine Learning?

## Definition

"A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."

-- Tom M. Mitchell, 1997

# What is Machine Learning?

## Definition

"A computer program is said to **learn** from experience E with respect to some class of tasks T and performance measure P if its performance at tasks in T, as measured by P, improves with experience E."
-- Tom M. Mitchell, 1997

Machine learning is the study of algorithms that

- improve their performance *P*
- at some task *T*
- with experience *E*

A well-defined learning task is given by *<P, T, E>*.

# What is Machine Learning?

Exercise: how to define the learning task in our Tea Example?

- improve their performance *P*
- at some task *T*
- with experience *E*

# Tea Example

# supervised learning

To learn the **mapping** (the rules) between a set of **inputs** and **outputs**.

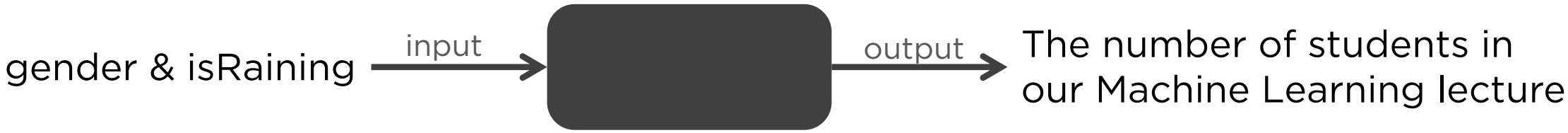
The tea example --


Goal: to learn the mapping that describes the relationship between
feature of tea (caffeine & acidity) and
type of tea (oolong tea / green tea).

**Labelled** data is provided of past input & output pairs during the learning process to train the model how it should behave.
So, **“supervised” learning**.

# What if the output is a continuous value?

An example --




Alternatively, the output could be a **real-world scalar** (output a number).

This is known as regression.

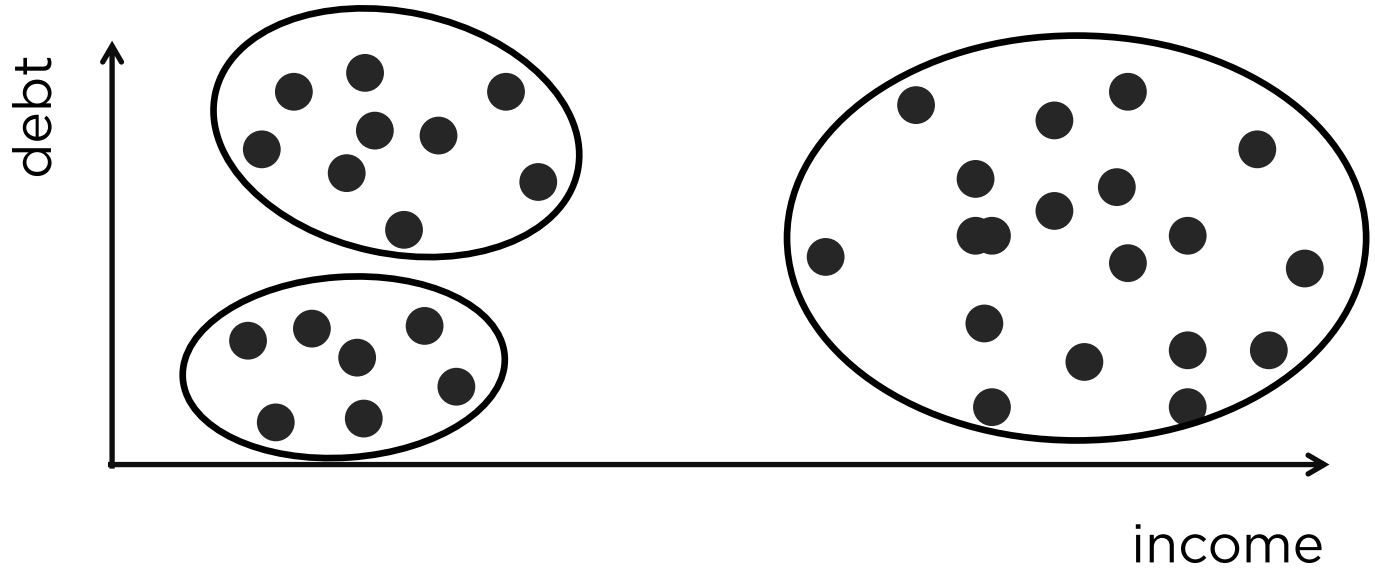
# unsupervised learning

To learn the **hidden pattern** (the rules) from a set of **inputs (no output)**.

**Unlabelled** data is provided of past input (not a input & output pair) during the learning process to train the model how it should behave.

So, **"unsupervised" learning**.

An example --




Goal: clustering a set of examples in such a way that the examples in the same group (called cluster) are more similar (in some sense or another) to each other than to those in other groups (clusters).

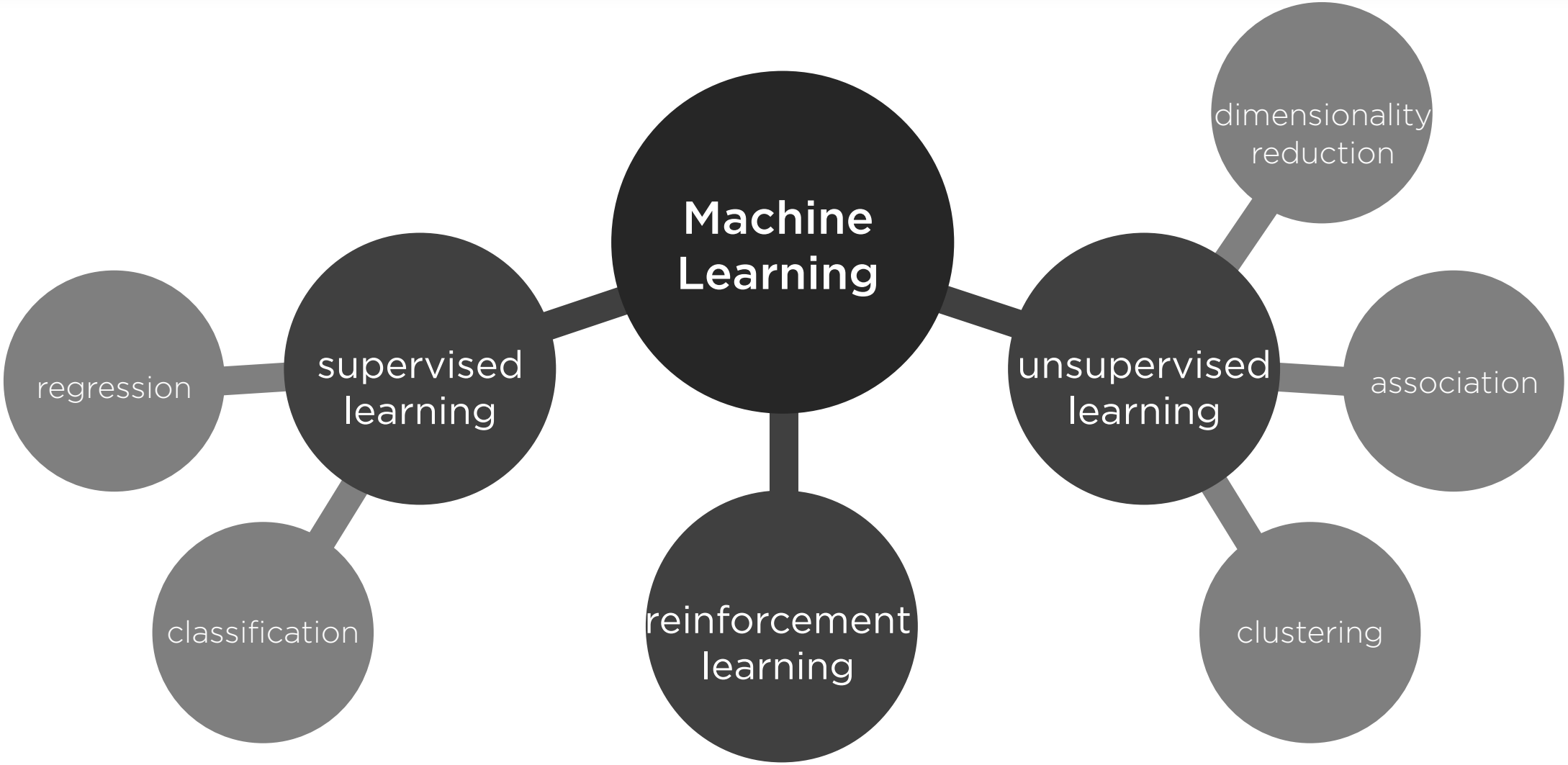
# reinforcement learning

Occasional **positive** and **negative** feedback is used to **reinforce** behaviours.

- **Good** behaviours are **rewarded** with a treat and become more common.
- **Bad** behaviours are **punished** and become less common.

A Reinforcement Learning algorithm just aims to maximise its rewards by playing the game over and over again.

# Three Categories of Machine Learning



# Key ML Terminology

# Key ML Terminology

- **Label** is the variable that we are predicting

  typically represented by the variable **y**
- **Features** are input variables that describe our data

  typically represented by the variables $\{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \ldots, \mathbf{x}_n\}$

# Key ML Terminology

- **Example** is a particular instance of data, $\mathbf{x}$
  - **Labelled example** has {features, label}: $(\mathbf{x}, \mathbf{y})$

    used to train the model
  - Unlabelled example has {feature, ?}: $(\mathbf{x}, \mathbf{?})$

    used for making prediction on new data

# Key ML Terminology

- **Model** maps examples to predict labels: $\hat{y}$

  defined by internal parameters, which are learned

  - **Training** - creating or learning the model.
  - **Inference** - applying the trained model to unlabelled examples.

# What to cover in the module?

## Lectures (tentative)

1. Introduction to Machine Learning
2. Linear Regression, Training and Loss
3. Generalisation, Training Test Set, Representation
4. Binary Classifier, Performance Measurement
5. Odds, Log(odds), Logistic Regression
6. Decision Trees and Forests
7. Support Vector Machines and Kernel Methods
8. Clustering
9. Neural Networks, Deep Learning, and Bias in AI
10. ML Applications

# General Module Information

# Contact



Dr SHI Lei


lei.shi@liverpool.ac.uk


0191 33 48131


http://shilei.me


E231

# Submodule Delivery

## Ten one-hour Lectures

Time: Mondays 12-1 PM, Weeks 11-20

Room: D110

## Five two-hour Labs

Time: Weeks 12, 14, 16, 18, 20

Room: E216A-B

# Assessment

## Assignment

Hand out: 31 January 2020

Hand in: 06 March 2020

Hand back: 04 May 2020

## Revision

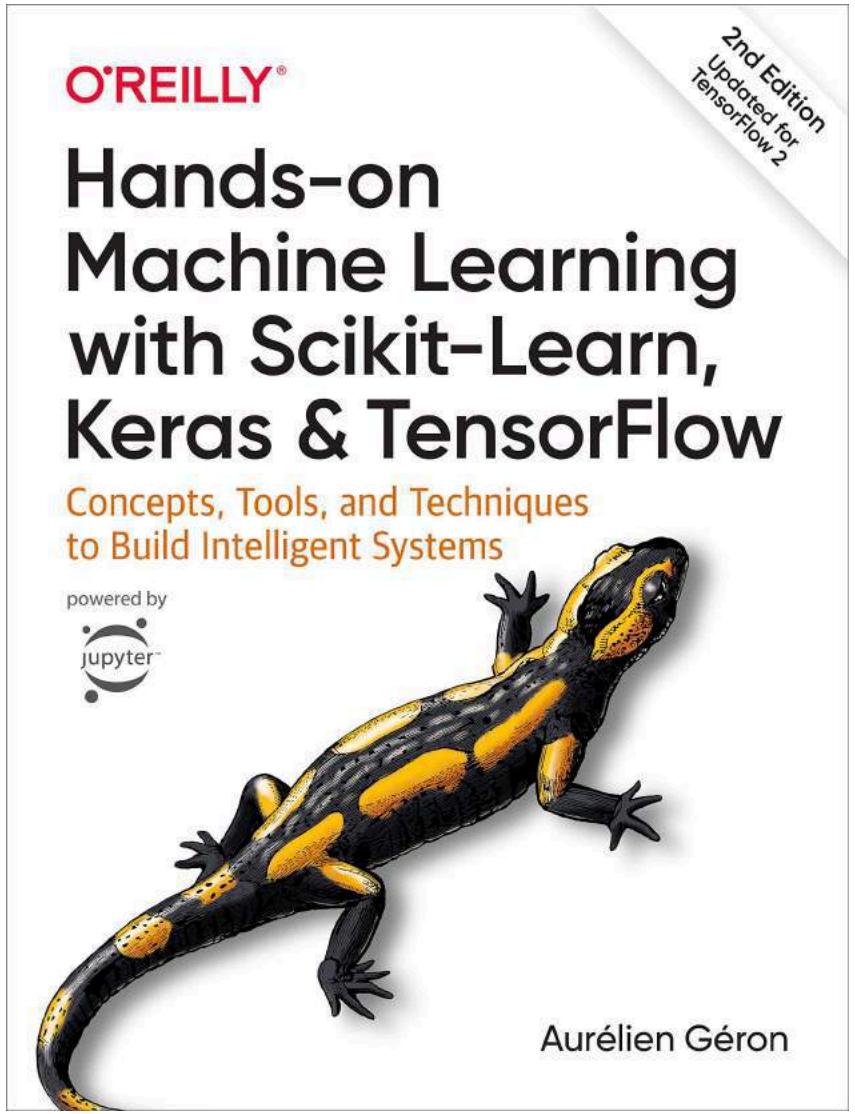
Time: 04 May 2020 12:00-12:45

Room: D110

## Exam

May 2020 (TBD)

# Reading & Practising




Aurélien Géron, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, 2nd Edition, O'Reilly Media, Inc. Sep 2019

# Summary

# Today

- What is Machine Learning (ML)
  - Definition
  - Lifecycle
  - Types of ML Systems
  - Key Terminologies
- General Module Information

# Homework

- On DUO – The Machine Learning Landscape

# Next Lecture

- Linear Regression
- Training and Loss