Week 3: Unsupervised Learning and Reinforcement Learning

Start with a question

Think first

How can a machine learn something useful if no correct answers are given?

Start with a question

Think first

How can a machine learn something useful if no correct answers are given?

Discussion prompt

What kinds of patterns could a model still discover in the data?

Unsupervised learning

Idea

Unsupervised learning uses data without labels.

Main goal

The goal is to find structure in the data.

- no correct answers are given
- the model looks for patterns, groups, or relationships
- useful when there is a lot of data to organise

Key phrase

Unsupervised learning means **learning without labels**.

Clustering as optimisation

Main idea

Minimise loss by adjusting cluster centres, keeping similar data points together.

$$L = \sum_{i=1}^{n} \|x_i - c_j\|^2$$

- x_i = data point
- c_j = assigned cluster centre

Interpretation

Minimise distances within groups.

K-means intuition

What the algorithm tries to do

- choose cluster centres
- assign each point to the nearest centre
- update the centres
- repeat

Geometric view

The data space is split into regions, one for each cluster.

Key idea

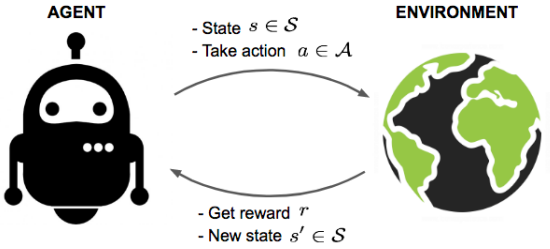
K-means is a simple example of optimisation in unsupervised learning.

Reinforcement learning

Idea

Reinforcement learning is learning by trying actions and receiving rewards.

- **agent**: the learner
- **environment**: what the agent interacts with
- **state** s: current situation
- **action** a: what the agent does
- **reward** r: feedback from the environment



Expectations in reinforcement learning

Why expectations matter

In RL, the next reward is often uncertain, so we talk about the *expected* outcome.

Expectations in reinforcement learning

Why expectations matter

In RL, the next reward is often uncertain, so we talk about the *expected* outcome.

Idea

Instead of asking what happens in one specific future, we ask what happens **on average**.

- a policy may lead to different outcomes each time
- the agent cares about the average long-term reward
- this is why RL often uses expected return or expected value

Key

Expectation helps us deal with uncertainty.

Markov chains as a simple model

Idea

A Markov chain describes a process where the next state depends only on the current state.

Markov chains as a simple model

Idea

A Markov chain describes a process where the next state depends only on the current state.

Robot leaving a circle

Imagine a robot wandering inside a circle. We approximate the circle with a small grid and let the robot move randomly between nearby cells.

- each cell is a **state**
- each move changes the state with some probability
- the boundary is an **exit** or **terminal** state

Key idea

A Markov chain is a useful approximation when we only care about the current position, not the full history.

Markov chains: grid examples

Boundary cells have value 0. Equal probability of moving up, down, left, or right. Fill in the interior values with the expected values if each move costs 1. Gray cells are walls and cannot be entered. How does this affect the probabilities?

Example 1

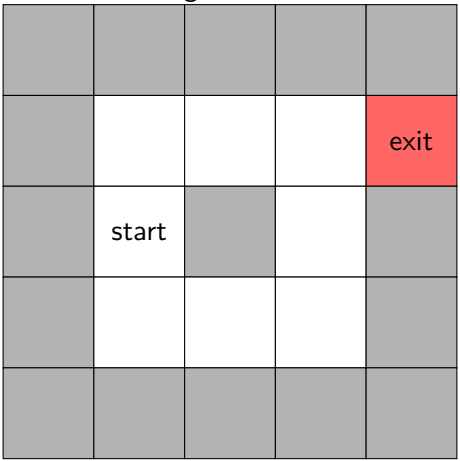
0	0	0	0
0	?	?	0
0	?	?	0
0	0	0	0

Example 2

0	0	0	0	0
0	?		?	0
0	?		?	0
0	?	?	?	0
0	0	0	0	0

What if the robot can choose where to go

Each move costs 1, leaving costs 0. Fill in the white boxes.



Policy and reward

Key idea

Reality is something in between random and controlled. Either way the agent learns which actions lead to better long-term rewards.

Policy

A policy tells the agent how to behave:

$$\pi(s) \rightarrow a$$

Objective

The goal is to maximise total reward:

$$\sum_{t=0}^{T} r_t$$

Examples and applications

Examples in practice

- K-means clustering for grouping similar data
- neural networks for more advanced pattern discovery
- separating large amounts of data into meaningful groups
- game-playing agents in reinforcement learning

Big picture

Unsupervised learning finds structure. Reinforcement learning finds better actions.

Recap

Today

- unsupervised learning finds structure without labels
- clustering can be written as an optimisation problem
- reinforcement learning learns by maximising reward
- policy, state, action, and reward are the main RL ideas
- For more, see this video from Google about how they made a Reinforcement Learning agent that beat a world champion: link to video

Next lesson

We will explain how ChatGPT and neural networks do what they do.