May 2025
Mikiyas Daniel

Running Behind

Exploring the Race Between AI Advancements and Our Understanding

Course Content

Course Outline:

1. Introduction – "Running Behind"
2. Purpose of the Course
3. AI and Mathematics
4. Algorithms in AI – Focus on Markov Chains
5. Demonstration with Example and Simulation with code
6. Conclusion – Bridging the Gap

Purpose of the Course

Why This Course?

- To **shed light on the vital role of mathematics**—especially **probabilistic models like Markov Chains**—in powering modern AI systems.
- To **explore how AI makes sequential decisions** and predictions over time using **Markov-based logic**, which models real-world behaviors such as language, movement, or user activity.
- To **understand why those without a strong mathematical foundation**—particularly in concepts like **state transitions, probability matrices, and stochastic processes(sequence of results)**—often fall behind in grasping the true mechanics of AI.
- To **ignite curiosity and deeper learning** in how **mathematical reasoning shapes intelligent behavior**, equipping learners to keep pace with AI's rapid evolution.

AI & Mathematics – The Hidden Engine

AI without Math is Like a Car without an Engine

- **Linear algebra**: for neural networks and deep learning.
- **Probability & statistics**: for predictions, uncertainty, and model evaluation.
- **Calculus**: for optimization and training models.
- **Discrete math & logic**: for decision-making systems.

Key Point: Every AI algorithm you see today runs on solid mathematical foundations.

What is a Markov Chain?

Markov Chain

Andrey Andreyevich Markov (14 June 1856 – 20 July 1922) was a Russian mathematician best known for his work on **stochastic processes**. A primary subject of his research later became known as the **Markov chain**.

There are four common Markov models used in different situations, depending on whether every sequential state is observable or not, and whether the system is to be adjusted on the basis of observations made:

> **“Our scope is under system is autonomous and system state is fully Observable”**

Ctd ...

A **Markov Chain** is a **mathematical model** for a system that transitions from one state to another, where:

> **The probability of moving to the next state depends *only* on the current state and not on the sequence of states that preceded it.**

This is called the **Markov Property**.

Ctd ...

Use in AI:

- Predictive text (e.g., keyboards)

  How AI Works: Suggests next words based on what you've typed—using **Markov chains** or **N-gram models** to predict likely word sequences.
  Math: Probabilities of word sequences are learned from large text corpora.

- Web page ranking (Google PageRank)

  -How AI Works: Google's PageRank algorithm uses a **Markov chain** to model the probability that a user will land on a certain webpage by randomly clicking links.
  -Math: States = web pages; Transitions = hyperlink clicks; Markov model helps assign rank based on long-term probabilities.

- Weather prediction
- Game development (NPC behavior)

Ctd …

A Markov chain is simplest type of Markov model, where all states are observable and probabilities converge over time.

- Describe the world in a more realistic way,
- Are a useful tool to make long-term predictions about a system or process.

Mathematical Illustration With Example:

Workout routine as a Markov Chain

Mathematical Illustration:

We pick one exercise at random and go from there but, the next set is not completely random, it usually depending on the exercise we did before.

So, in a typical workout day, you end up doing several sets of these exercises:

- Run (3 miles)
- Push-ups (20)
- Rowing (15 minutes)
- Pull-ups (20)

Ctd…

‘’A Markov chain has short-term memory, it only remembers where you are now and where you want to go next.”

This means the path you took to reach a particular state doesn’t impact the likelihood of moving to another state. The only thing that can influence the likelihood of going from one state to the other is the state you are currently in.

We want to understand more about our optimal **workout routine** and even plan the next workout based on how you are normally structure it. So we realize that your workout routine can be modeled as a **Markov chain.**

Ctd ..

Knowing all of this, you can ask interesting questions about your workout:

- If I start the workout with a run, how likely am I to do push-ups on the second set?
- In a 3-set workout, how likely am I to do: 1) run, 2) push-ups and 3) pull-ups?
- What’s the likelihood of doing any of the exercises on the fourth set?

Ctd ..

Thankfully lets say we keep a good record of your workout sets and looking at the **last 62 days**, we had the following routine:

Set	Run (3 miles)	Push-ups (20)	Rowing (15 minutes)	Pull-ups (20)	Total
Run (3 miles)	1	4	3	2	10
Push-ups (20)	7	2	5	6	20
Rowing (15 minutes)	8	6	1	5	20
Pull-ups (20)	5	5	1	1	12
	21	17	10	14	62

Ctd ..

-Considering each exercise set as a state in your workout Markov Chain, the next thing you do is to encode the dependencies between states, using **conditional probabilities.**

-In the context of Markov models, these conditional probabilities are called **transition probabilities.**

Transition probabilities describe the transition between states in the chain as a conditional probability.

Ctd ..

After crunching the numbers your transition probabilities are:

	P(Next set | Set)				
Set	Run (3 miles)	Push-ups (20)	Rowing (15 minutes)	Pull-ups (20)	Total
Run (3 miles)	0.10	0.40	0.30	0.20	1.00
Push-ups (20)	0.35	0.10	0.25	0.30	1.00
Rowing (15 minutes)	0.40	0.30	0.05	0.25	1.00
Pull-ups (20)	0.42	0.42	0.08	0.08	1.00

Transition probabilities for each sequence of exercise sets.

So, the likelihood of doing 20 push-ups *given that* you've just finished a 3 mile run is 40%. That's because P(20 push-ups | Run) = 4/10 = 0.4.

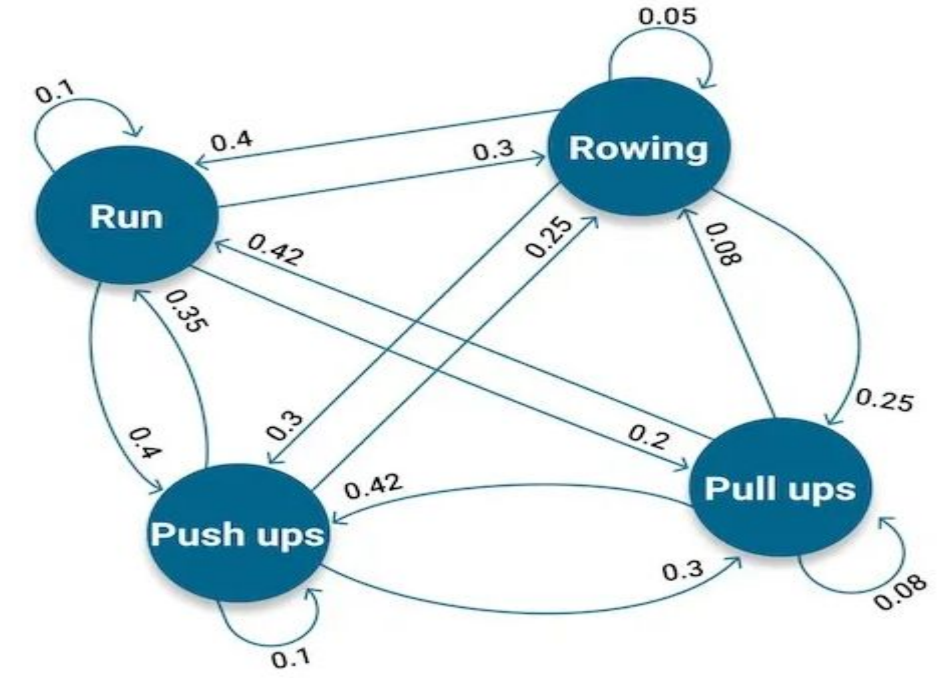
Ctd ..

Put in mathematical notation, these probabilities can be represented as a **transition matrix**[4].

$$
P = \begin{array}{r} \\ \text{Run} \\ \text{Push-ups} \\ \text{Rowing} \\ \text{Pull-ups} \end{array}
\begin{array}{c} \begin{array}{cccc} \text{Run} & \text{Push-ups} & \text{Rowing} & \text{Pull-ups} \end{array} \\
\begin{pmatrix} 0.10 & 0.40 & 0.30 & 0.20 \\ 0.35 & 0.10 & 0.25 & 0.30 \\ 0.40 & 0.30 & 0.05 & 0.25 \\ 0.42 & 0.42 & 0.08 & 0.08 \end{pmatrix} \end{array}
$$

Ctd ..

And to better visualize the transitions between states, you can represent the workout Markov Chain as a directed graph.



Your workout routine modeled as a Markov chain.

Example:-

If I start the workout with a run, how likely am I to do push-up after 2 steps ?

To understand how the workout routine evolves over time there are three components to take into account:

- Starting state,
- End state,
- Time-frame, i.e., how long it will take the model to get from the start to the end state.

Ctd ..

If I start the workout with a run, how likely am I to do push-ups on the Second set?

-Since we're making predictions for the third set of exercises, the time-frame is two.

Start (1) , Second set (2).

Next we'll calculate the *total probability*. We'll combine all the possible ways, or paths in the Markov chain, where we start the workout with a run and in two time steps do push-ups.

Given this criteria, you have the following paths:

Ctd ...

Path 1: Go for run given that you've just completed a run, then do a set of push-ups.

$$p_{\text{Run,Run}} \cdot p_{\text{Push-ups,Run}}$$

Path 2: Do 20 push-ups given that you've just completed a run, then do another set of push-ups.

$$p_{\text{Push-ups,Run}} \cdot p_{\text{Push-ups,Push-ups}}$$

Path 3: Row for 15 minutes given that you've just completed a run, then do 20 push-ups.

$$p_{\text{Rowing,Run}} \cdot p_{\text{Push-ups,Rowing}}$$

Ctd …

Path 4: Do 20 pull-ups given that you've just completed a run, then do 20 push-ups.

$$p_{\text{Pull-ups,Run}} \cdot p_{\text{Push-ups,Pull-ups}}$$

where you start with a run and end with push-ups is 25%

$$P^{(2)}_{Run,\ Push-ups} = p_{\text{Run,Run}} \cdot p_{\text{Push-ups,Run}} + p_{\text{Push-ups,Run}} \cdot p_{\text{Push-ups,Push-ups}} + p_{\text{Rowing,Run}} \cdot p_{\text{Push-ups,Rowing}} + p_{\text{Pull-ups,Run}} \cdot p_{\text{Push-ups,Pull-ups}}$$

$$\equiv \left[P_{(\text{Run} \mid \text{Run})} . P_{(\text{Push-ups} \mid \text{Run})}\right] + \left[P_{(\text{Push-ups} \mid \text{Run})} . P_{(\text{Push-ups} \mid \text{Push-ups})}\right] + \left[P_{(\text{Rowing} \mid \text{Run})} . P_{(\text{Push-ups} \mid \text{Rowing})}\right] + \left[P_{(\text{Pull-ups} \mid \text{Run})} . P_{(\text{Push-ups} \mid \text{Push-ups})}\right]$$

$$\equiv (0.1 \times 0.4) + (0.4 \times 0.1) + (0.3 \times 0.3) + (0.2 \times 0.42)$$

$$\approx 0.04 + 0.04 + 0.09 + 0.0833$$
$$\approx 0.253$$

Ctd …

Putting it all these paths together, the likelihood of doing a two-set**(two time step)** workout where you start with a run and end with push-ups is 25%.

Adopting a matrix notation you can extend the probabilities of the two-set workout to all other combinations of **start and end states**.

Since this corresponds to what the chain will look like in two time-steps, this matrix called the **second power of the transition matrix**.

Ctd …

Adopting a matrix notation you can extend the probabilities of the two-set

$P^2 =$

	Run	Push-ups	Rowing	Pull-ups
Run	0.35	0.25	0.16	0.23
Push-ups	0.30	0.35	0.17	0.19
Rowing	0.28	0.31	0.22	0.2
Pull-ups	0.26	0.27	0.24	0.24

Markov Chain General Formula

For a discrete-time Markov chain, the **transition probability** is given by:

$$P(X_{n+1}=j \mid X_n=i)=P_{ij}$$

-X_n is the state at time step n

-i, j are specific states

-P_{ij} is the probability of moving from state i to state j

To predict where the system will be in **n steps**, use:

$$P^{(n)} = P^n$$

Drawbacks of Markov Chains

1. Markov Property Limitation (Memoryless)

"The future depends only on the present, not on the past."

- Problem : Markov Chains assume that the next state depends only on the current state , not on the sequence of events that led to it.
- Limitation : This can be too simplistic for real-world scenarios where history matters .

  📌 Example:

  - **In fitness tracking:** If someone is running, knowing whether they just did push ups or rested might affect what they'll do next — but a basic Markov Chain ignores this.

Ctd ..

2. Difficulty in Modeling Complex Dependencies

- Since the chain has no memory, it cannot naturally model long-term dependencies or context-sensitive transitions .
- You may need to artificially expand the state space (e.g., using higher-order Markov chains), which increases complexity.

3. Stationary Distribution May Not Be Reached Fast Enough

-While many Markov Chains converge to a stationary distribution, some converge very slowly .

-In practice, if you're simulating a system, you might not have enough steps/time to reach that stable behavior.

Conclusion – Catching Up

Key Takeaways:

- We are "Running Behind" not just in **technology use**, but in understanding **how it works**.
- To truly grasp and build AI systems, **mathematics is essential**.
- The future belongs to those who can blend **math thinking with AI innovation**.
- Start small: one algorithm, one formula, one step at a time

THE END

Q & A