# **Introduction**

In this session, you'll learn the most basic strategy in reinforcement learning (****Dynamic Programming)****to compute optimal policies ****given the model of the environment****. You'll learn that there are two basic steps for solving any RL problem.

* ****Policy evaluation**** refers to the iterative computation of the value functions for a given policy.
* ****Policy improvement**** refers to finding an improved policy given the value function for an existing policy.

You'll learn about two basic approaches to solve an RL problem:

* ****Policy Iteration**** and
* ****Value Iteration****

Both these techniques use these two steps repeatedly to arrive at optimality. Either of these can be used to reliably compute optimal policies and value functions given complete knowledge of the MDP.

You'll also learn to apply RL in an Ad Placement Optimisation Demo.

## Prerequisites

There are no prerequisites for this session.

## Guidelines for In-Module Questions

The in-video and in-content questions for this module are not graded. Note that graded questions are given on a separate page labelled 'Graded Questions' at the end of this session. The graded questions in this session will adhere to the following guidelines:

|  |  |  |  |
| --- | --- | --- | --- |
|  | First Attempt Marks | |  | | --- | | Second Attempt Marks | |
| Questions with 2 Attempts | 10 | 5 |
| Questions with 1 Attempt | 10 | 0 |

## People you will hear from in this session:

****Subject Matter Expert:****

[G. Srinivasaraghavan](https://www.linkedin.com/in/gopalakrishnan-srinivasaraghavan-43b4b9?authType=NAME_SEARCH&authToken=uGg0&locale=en_US&trk=tyah&trkInfo=clickedVertical:mynetwork,clickedEntityId:1239599,authType:NAME_SEARCH,idx:1-1-1,tarId:1476708729573,tas:Srinivasaraghavan iiit )

Professor, IIIT-Bangalore

The International Institute of Information Technology, Bangalore, commonly known as IIIT Bangalore, is a premier national graduate school in India. Founded in 1999, it offers Integrated M.Tech., M.Tech., M.S. (Research) and PhD programs in the field of Information Technology.

# **Dynamic Programming**

****Dynamic Programming (DP)**** is a method of solving complex problems by breaking them into sub-problems. It solves the sub-problems and stores the solution of sub-problems, so when next time same subproblem arises, one can simply look up the previously computed solution.

Why should you apply dynamic programming to solve Markov Decision Process?

1. ****Bellman equations are recursive in nature****. You can break down Bellman equation to two parts – (i) optimal behaviour at next step (ii) optimal behaviour after one step
2. You can ****cache the state value and action value functions**** for further decisions or calculations.

The basic assumption in this session is that:****model of the environment****p(s′,r|s,a)is available.

## Additional Reading

You can read more on Dynamic Programming [here](https://web.stanford.edu/class/cs97si/04-dynamic-programming.pdf" \t "https://learn.upgrad.com/course/1688/segment/14428/89172/267129/_blank).