# CS 747, Autumn 2022: Lecture 0

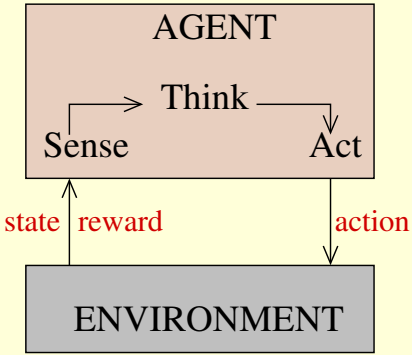
Shivaram Kalyanakrishnan

Department of Computer Science and Engineering
Indian Institute of Technology Bombay

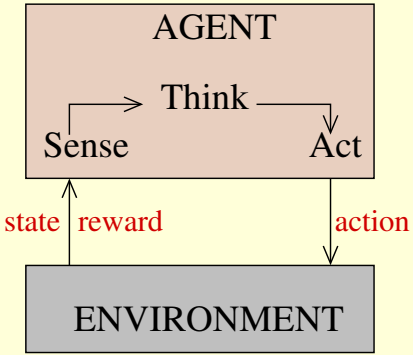
Autumn 2022

# CS 747: Foundations of Intelligent and Learning Agents

# CS 747: Foundations of Intelligent and Learning Agents

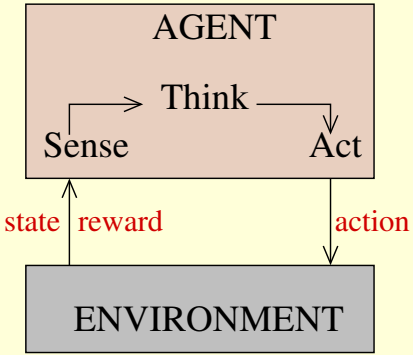


# CS 747: Foundations of Intelligent and Learning Agents



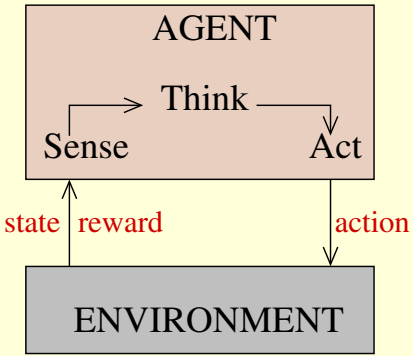

| Agent | Environment/Task |
|---|---|
| Algorithm for . . . | Stock trading |

# CS 747: Foundations of Intelligent and Learning Agents



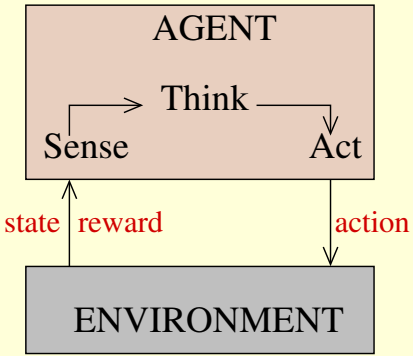

| Agent | Environment/Task |
|---|---|
| Algorithm for . . . | Stock trading |
| Autopilot program | Airplane |

# CS 747: Foundations of Intelligent and Learning Agents



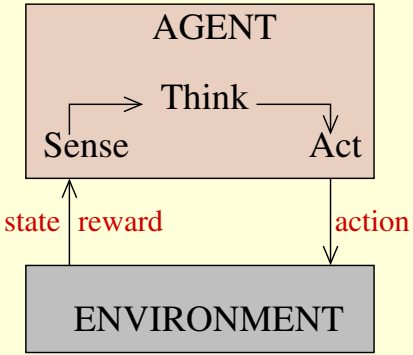

| Agent | Environment/Task |
|---|---|
| Algorithm for . . . | Stock trading |
| Autopilot program | Airplane |
| You | Bicycle |

# CS 747: Foundations of Intelligent and Learning Agents




| Agent | Environment/Task |
|---|---|
| Algorithm for . . . | Stock trading |
| Autopilot program | Airplane |
| You | Bicycle |
| You | Academic programme at IIT Bombay |

# CS 747: Foundations of Intelligent and Learning Agents




| **Agent** | **Environment/Task** |
|---|---|
| Algorithm for . . . | Stock trading |
| Autopilot program | Airplane |
| You | Bicycle |
| You | Academic programme at IIT Bombay |
| AlphaGo | Go |

# CS 747: Foundations of Intelligent and Learning Agents

- **Multi-armed bandits**
- **Markov Decision Problems**
- **Reinforcement learning**
- **Multi-agent systems/learning**

# CS 747: Foundations of Intelligent and Learning Agents

- **Multi-armed bandits**
  - The “explore or exploit” tradeoff.
- **Markov Decision Problems**
- **Reinforcement learning**
- **Multi-agent systems/learning**

# CS 747: Foundations of Intelligent and Learning Agents

- **Multi-armed bandits**
  - The “explore or exploit” tradeoff.
- **Markov Decision Problems**
  - Sequential decision making.
- **Reinforcement learning**

- **Multi-agent systems/learning**

# CS 747: Foundations of Intelligent and Learning Agents

- **Multi-armed bandits**
  - The “explore or exploit” tradeoff.
- **Markov Decision Problems**
  - Sequential decision making.
- **Reinforcement learning**
  - Learning by trial and error, reward and punishment, to optimise long-term gain.
- **Multi-agent systems/learning**

# CS 747: Foundations of Intelligent and Learning Agents

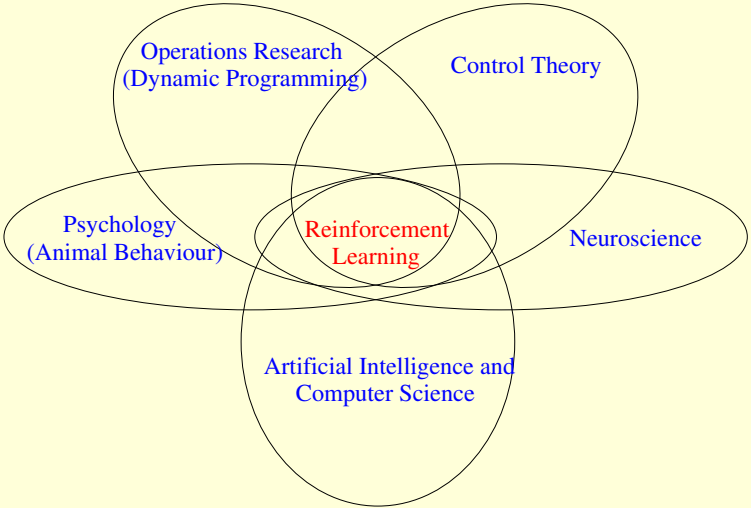
- **Multi-armed bandits**
  - The “explore or exploit” tradeoff.
- **Markov Decision Problems**
  - Sequential decision making.
- **Reinforcement learning**
  - Learning by trial and error, reward and punishment, to optimise long-term gain.
- **Multi-agent systems/learning**
  - Decision making in the presence of other decision-makers.

# CS 747: Foundations of Intelligent and Learning Agents

- **Multi-armed bandits**
  - The “explore or exploit” tradeoff.
- **Markov Decision Problems**
  - Sequential decision making.
- **Reinforcement learning**
  - Learning by trial and error, reward and punishment, to optimise long-term gain.
- **Multi-agent systems/learning**
  - Decision making in the presence of other decision-makers.

Several applications: game playing, robotics and control, planning and scheduling, on-line advertising, autonomous navigation, chemistry!

# CS 747: Foundations of Intelligent and Learning Agents



# CS 747: Foundations of Intelligent and Learning Agents

- **Prerequisites**: Probability, Algorithms, Programming.

# CS 747: Foundations of Intelligent and Learning Agents

- **Prerequisites**: Probability, Algorithms, Programming.
- Coming up in two weeks!:

$$
\begin{aligned}
B &\leq \sum_{t=0}^{T-1} \sum_{x=\widetilde{u}_a^T}^{t} \sum_{y=1}^{t} \mathbb{P}\left\{ \hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)} \right\} \\
&\leq \sum_{t=0}^{T-1} \sum_{x=\widetilde{u}_a^T}^{t} \sum_{y=1}^{t} \left( \mathbb{P}\left\{ \hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2} \right\} + \mathbb{P}\left\{ \hat{p}_\star(y) < p_\star - \sqrt{\frac{2}{y}\ln(t)} \right\} \right) \\
&\leq \sum_{t=0}^{T-1} \sum_{x=\widetilde{u}_a^T}^{t} \sum_{y=1}^{t} \left( e^{-2x\left(\frac{\Delta_a}{2}\right)^2} + e^{-2y\left(\sqrt{\frac{2}{y}\ln(t)}\right)^2} \right) \\
&\leq \sum_{t=0}^{T-1} \sum_{x=\widetilde{u}_a^T}^{t} \sum_{y=1}^{t} \left( e^{-4\ln(t)} + e^{-4\ln(t)} \right) \leq \sum_{t=0}^{T-1} t^2 \left( \frac{2}{t^4} \right) \leq \sum_{t=0}^{\infty} \frac{2}{t^2} = \frac{\pi^2}{3}.
\end{aligned}
$$

# CS 747: Foundations of Intelligent and Learning Agents

- **Prerequisites**: Probability, Algorithms, Programming.
- Coming up in two weeks!:

$$
\begin{aligned}
B &\leq \sum_{t=0}^{T-1} \sum_{x=\bar{v}_a^T}^{t} \sum_{y=1}^{t} \mathbb{P}\left\{ \hat{p}_a(x) + \sqrt{\frac{2}{x}\ln(t)} \geq \hat{p}_\star(y) + \sqrt{\frac{2}{y}\ln(t)} \right\} \\
&\leq \sum_{t=0}^{T-1} \sum_{x=\bar{v}_a^T}^{t} \sum_{y=1}^{t} \left( \mathbb{P}\left\{ \hat{p}_a(x) \geq p_a + \frac{\Delta_a}{2} \right\} + \mathbb{P}\left\{ \hat{p}_\star(y) < p_\star - \sqrt{\frac{2}{y}\ln(t)} \right\} \right) \\
&\leq \sum_{t=0}^{T-1} \sum_{x=\bar{v}_a^T}^{t} \sum_{y=1}^{t} \left( e^{-2x\left(\frac{\Delta_a}{2}\right)^2} + e^{-2y\left(\sqrt{\frac{2}{y}\ln(t)}\right)^2} \right) \\
&\leq \sum_{t=0}^{T-1} \sum_{x=\bar{v}_a^T}^{t} \sum_{y=1}^{t} \left( e^{-4\ln(t)} + e^{-4\ln(t)} \right) \leq \sum_{t=0}^{T-1} t^2 \left( \frac{2}{t^4} \right) \leq \sum_{t=0}^{\infty} \frac{2}{t^2} = \frac{\pi^2}{3}.
\end{aligned}
$$

- Video lectures, solved quizzes and exams, programming assignments all linked from previous years' course pages.

  `https://www.cse.iitb.ac.in/~shivaram/teaching/old/cs747-a2021/index.html`