**Vivekanand Education Society’s Institute of Technology**

**Department of AIDS Engineering**



**Subject:**

**Class: D16AD**

| **Roll No: 30** | **Name:** [**SUHANEE KANDALKAR**](mailto:2021.suhanee.kandalkar@ves.ac.in) |
| --- | --- |
| **Practical No: 8** | **Title: Double Q learning** |
| **DOP:** | **DOS:** |
| **Grades:** | **LOs Mapped: LO** |
| **Signature:** |  |

**Double Q-Learning :**

Double Q-Learning is an improvement over traditional Q-Learning that addresses the issue of maximization bias—a problem where Q-Learning tends to overestimate action values due to always selecting the maximum estimated reward. This overestimation can lead to suboptimal decision-making, especially in complex environments.

In Double Q-Learning, two separate Q-tables (Q1 and Q2) are maintained. The key idea is that:

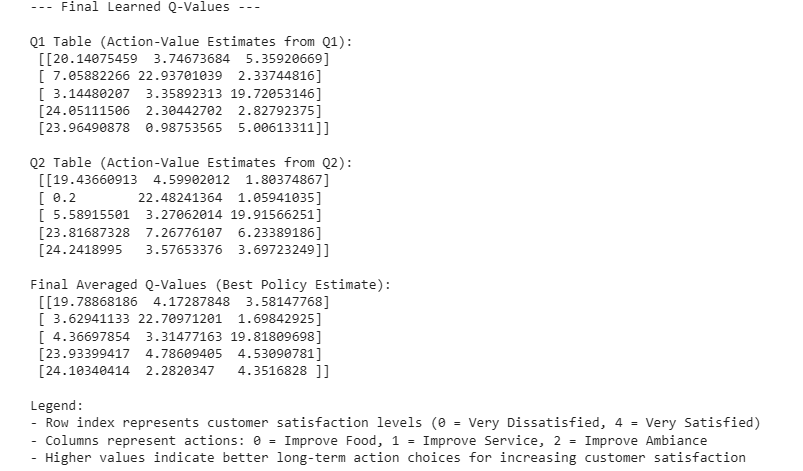
* One table (Q1) is used to select the best action for the next state.
* The other table (Q2) is used to evaluate that action’s value.
* The process is randomly alternated between the two tables, reducing the likelihood of overestimation.

By splitting action selection and evaluation between these two tables, Double Q-Learning provides more stable learning and leads to better policy convergence.

In the given experiment, we simulate a restaurant trying to improve customer satisfaction using reinforcement learning. The states represent different levels of customer satisfaction, ranging from very dissatisfied to very satisfied. The actions available to the restaurant include improving food, improving service, or enhancing ambiance .Using Double Q-Learning, the restaurant learns which improvements lead to the highest long-term satisfaction. Instead of relying on a single Q-table (which might overestimate the impact of an action), two Q-tables (Q1 and Q2) work together to refine action value estimates.For example, if the restaurant is currently at satisfaction level 2 (Neutral) and decides to improve service, it might move to level 4 (Very Satisfied), receiving a high reward. Over time, the algorithm learns that improving service generally provides the most significant boost to satisfaction.By alternating between Q1 and Q2 for updates, the restaurant avoids overvaluing specific actions and instead discovers the most balanced and effective strategy for long-term customer satisfaction improvement.

**Code :**

[**EXP\_8\_RL\_30.ipynb**](https://colab.research.google.com/drive/1B7EazoHnCEsJepAl_ghOnB2O_9Xsxw8x?usp=sharing)

****

**Conclusion:**

Double Q-Learning effectively reduces maximization bias by maintaining two Q-tables (Q1 and Q2) that alternate action selection and evaluation. This results in more stable learning and better policy decisions.

In the restaurant rating scenario, the model learns which actions—improving food, service, or ambiance—lead to the highest long-term customer satisfaction. Unlike standard Q-Learning, Double Q-Learning prevents overestimation, ensuring a balanced and optimal strategy for improvement.This approach proves valuable for real-world decision-making in fields like business, robotics, and AI, where accurate long-term planning is crucial.