Week_8_df
In this unit you are introduced to reinforcement learning. Using your own words, compare the **planning** and **learning** algorithms. In your initial post, be sure to include:

- the mathematical concepts/reasoning used.

Mathematically Reinforcement Learning is making decisions by iterations, like the simplex method in linear programming and control problems.
Bellman equation correlates state-value function with other states. It is used for the computation value of a given policy. In addition, our goal is to find optimal policy and value function. CMT (Contraction Mapping Theorem) gives us a guarantee of policy evaluation.

This week we were learning about MDP or Markov Decision Process. One of the main problems is optimization. Two approaches are: reinforcement learning and planning.
Our environment model can be either known to the learner(agent) or unknown. In the case of a known environment model, the objective here is to make maximum reward for planning problems. Otherwise, a learner deals with a learning problem.
One of the most challenging problems of a learner in MDP is to “understand” what action to take based on the state. This is also called a policy. An agent intends to find a policy which maximizes its reward. As we can see here something interesting. A learner wants to get maximum of its policy value, thus we should somehow find a way to optimize his searching. But good for us, we know that for any state exists optimal policy.

- examples of its working domain and applications.

Example of planning could be a planning of a trip from the house to some destination. Such a process contains steps or states, constraints like estimated minimal time, or cheapest cost or most luxurious adventure.
Example of learning on the other hand could be an iRobot, which in its first steps tries to “learn” a house’ plan.

- Is one algorithm better to use for optimization problems? Explain your answer.

As obvious one word answer could be it depends. So, if our domain is when a model is known for the agent then the planning algorithm is better. On the other hand, when the environment model is unknown, the learning algorithm will shine.

We can discuss algorithms for each approach. For example, SARSA is a good learning algorithm when an agent navigates in an unknown environment. The number of states is not big. However, what if a number of states is big, how big? For instance 2^20. Then algorithm called TD(λ) is suitable better.

References:

1. https://arxiv.org/pdf/2006.15009.pdf
2. https://link.springer.com/article/10.1007/s10458-022-09552-y
3. https://arxiv.org/pdf/2107.07373.pdf
4. https://cims.nyu.edu/~donev/Teaching/WrittenOral/Projects/XintianHan-WrittenAndOral.pdf