FC Bias Variance Tradeoff

lschmid

May 2023

1 A lazy Sunday problem

Let's assume you are taking a lazy Sunday. Naturally in RL we model this as a MDP: There are two states: The first is doing chores, let's say cleaning, doing the dishes, calling your weird annoying family member because you have to, etc. Whenever you decide for doing chores you get a reward of zero.

Then there is the second state of doing something fun, having brunch with your friends, reading a good book, playing your favourite game, spending time with your loved ones (it's a cat and we all accept it!). Whenever you decide to do something fun, you get a reward of one. Whenever you are in either of the two states, you probably keep doing the same thing for some more time: Your policy is to keep doing the same thing with a probability of 0.95 and switching task with a probability of 0.05. You start the day doing one of the two things at random (i.e. with equal probability in either state)

1.1 Task 01

In this and the following tasks we will only calculate the expected Return for the starting state!
What is the expected value fo the Return of your lazy Sunday depending on γ? Tip: This is easy to compute when you keep in mind both states are equally likely at any given timestep, and how to compute the convergence of a Geometric Series. What is the expected value of the return over the first n steps?

1.2 Task 02

Let's say you take 36 (12h, one every 20 minutes) such decisions over a lazy Sunday, i.e. 36 steps.

- Write a function, which computes a MC-Return, given the starting state, *gamma*, and number of steps to compute it for (with states according to the policy and MDP above).
- Draw 500 such lazy Sundays with 36 timesteps, random starting state, and *gamma* as 0.95

- After each draw (i.e. for each of the 500 steps), compute the current MC estimate
- Plot it! (number of samples vs. current estimation)

1.3 Task 03

- Repeat the task above 500 times
- Plot the average MC-Estimate and its *std* for all sample numbers

1.4 Task 04

Compute the expected Return using the following simple TD-approach, where we estimate the expected return according to $r + \gamma v'$. Calculate the true v'. Assume different bias values (e.g. 0.01, 0.1, 0.2) in the following, which should be applied to it.

- Write a function, which computes the SARSA Estimate given the starting state, *gamma*, true Q-value, and bias to use for your Q-value.
- Draw 500 lazy sundays
- After each draw (i.e. for each of the 500 steps), compute the current average TD-Estimation
- Plot it! (number of samples vs. current estimation)

1.5 05

- Repeat the task above 500 times
- Plot the average TD-Estimate and its *std* for all sample numbers
- Plot both the TD-Estimate + *std*, and MC-Estimate + *std* in the same plot. Compare.