National University of Singapore
School of Computing

Semester 1, AY2021-22

CS4246/CS5446 AI Planning and Decision Making

Issued: 15 Oct, 2021

Tutorial Week 10: Function Approximation

Guidelines

You may discuss the content of the questions with your classmates. But everyone should work on and be ready to present ALL the solutions.

Problem 1: Review

Briefly explain the following concepts:

(a) Linear regression and loss function optimization.

(b) Gradient descent.

(c) Deep learning in neural networks.

Problem 2: Approximating TD Learning

[RN 3e 21.4] Write out the parameter update equations for temporal difference (TD) learning with

$$\hat{U}(x, y) = \theta_0 + \theta_1 x + \theta_2 y + \theta_3 \sqrt{(x - x_g) + (y - y_g)}.$$

Problem 3: Approximating Q-Learning

Consider a system with a single state variable x that can take value 0 or 1 and actions a_1 and a_2. An agent can observe the value of the state variable as well as the reward in the observed state. Assume a discount factor $\gamma = 0.9$.

(a) Perform two steps of Q-learning with the observed transitions shown below in (i) and (ii) using a table representation of the Q-function. Use a learning rate of $\alpha = 0.5$ starting from a table with all entries initialized to 0. Show the Q-function after each step.

(i) First observed transition: initial value of $x = 0$, observed reward $r = 10$, action a_1, next state $x = 1$.

(ii) Second observed transition: from $x = 1$, observed reward $r = -5$, action a_2, next state $x = 0$.

(b) Now perform Q-learning with function approximation using $Q(x, a_1) = \beta_1 x$ and $Q(x, a_2) = \beta_2 x$. Use a learning rate $\alpha = 0.5$ starting from parameters $\beta_1 = 0$ and $\beta_2 = 0$. Show the parameter values after each step.

(i) First observed transition: initial value of $x = 1$, observed reward $r = 10$, action a_1, next state $x = 1$.

(ii) Second observed transition: from $x = 1$, observed reward $r = -5$, action a_2, next state $x = 0$.

(c) After enough data is observed, which method would give better performance, the tabular method in (a) or the function approximation method in (b)? Why? Suggest how the poorer performing method can be improved.