b. Describe the TD(0) algorithm for policy prediction. Why, in general, does this compute a different solution from the Monte Carlo estimate?[10 marks]

TD(0) only looks one step ahead and can converge to some optimal solution given a fixed policy and a constrained step size. Monte Carlo methods minimize the MSE on training data whereas the TD(0) optimizes for the maximum-likelhood model of the Markov process. While both are optimal, they optimize for different conditions. TD(0) estimates the structureof the FMDP (or essentially, estimates a model) and subsequently performs policy evaluation on said model.

