Problem 1 (20)
(a) Explain the role of the feature vector, $F_a$, used by the Linear Gradient Descent Sarsa($\lambda$).
(b) Planning pertains to the use of a model for improving the learning process. Describe in what scenarios it proves beneficial to employ planning, and when can it potentially degrade the overall performance of the agent.

Problem 2 (15)
Why did the Dyna agent with exploration bonus, Dyna-Q+, perform better in the first phase as well as in the second phase of the blocking and shortcut experiments?

Problem 3 (15)
Careful inspection of Figure 9.8 (in the SB book) reveals that the difference between Dyna-Q+ and Dyna-Q narrowed slightly over the first part of the experiment. What might be the explanation for this?

Problem 4 (15)
Answer the following questions pertaining to function approximation methods:
1. What is the motivation for employing function approximation techniques in RL?
2. In what cases does gradient descent yield sub-optimal performance?
3. How would you reproduce the tabular case within the linear function approximation framework? In other words, what is the linear transformation that would yield the tabular case?

Problem 5 (15)
In accordance with the state diagram below, find the state-values, $V(i)$, for $i=1,2,3,4$. Assume that state 5 is a terminal state (with zero reward) and a discount factor $\gamma = 0.8$. 
Problem 6 (20)

In the forward view of TD($\lambda$), the parameter $\lambda$ characterizes how fast the exponential weighting falls off, and thus how far into the future the $\lambda$ return algorithm looks in determining its backup. However, a rate factor such as $\lambda$ is sometimes an awkward way of characterizing the speed of the decay. Alternatively, we may construct a function that simply records the last $n$ states visited and only marks those as eligible for updates. In other words, we have a finite window of $n$ time steps for determining the set of eligible update states (or state-action pairs). Formulate this solution in a similar manner to that presented for the exponential weighting scheme.