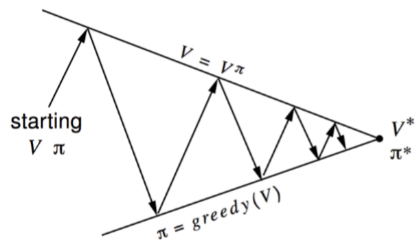
增强学习测试题

## 简答题（共10题，每题7分）

1. 简述如下概念：一阶马尔科夫属性、马尔科夫过程、马尔科夫回报过程、马尔科夫决策过程。
2. 列出机器学习领域的三个分支。
3. 简述增强学习适用于解决的问题的特点。
4. 列出将value function写成贝尔曼方程形式的两个好处。
5. 根据下图所示游戏，写出agent, state, reward分别是什么。

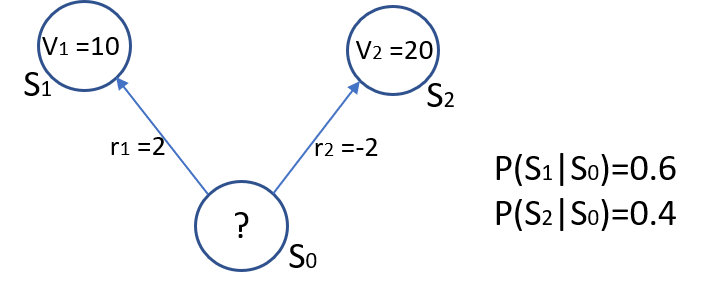


1. 根据下图，简述在model已知的情况下，使用policy iteration法求解最优policy的过程。



1. 在上问中，若是model未知的情况下，可以用MC或TD算法套用同样的过程，但是要做两个改动，分别是什么？
2. 从episode采样的利用、variance、bias这三个角度，比较蒙特卡洛算法和TD算法。
3. 简述TD(λ)算法的思想。
4. 如何理解on-policy和off-policy？

## 应用题（共3题，每题10分）

1. 如下图所示，点中V值代表该状态的v函数值，边上r值代表状态改变的立即回报，如果衰减因子gamma=0.9，请根据贝尔曼期望方程（已给出），计算？处的V值。
2. 若J(w)是关于w的一个可微分目标方程，若要使用梯度下降法更新w使J(w)的值减小，应该如何确定w的更新方向？
3. 若S是一个连续的无限的状态空间，已知这个空间中的状态最小值和最大值（Smax, Smin）， 如何将S 映射到[0,10]之间的整数。（提示：min-max标准化方法）