Bellman Equation for Policy Iteration

The Bellman equation for policy iteration is given by:

$$V^{\pi}(s) = \sum_{a} \pi(a|s) \sum_{s',r} p(s', r|s, a)[r + \gamma V^{\pi}(s')] \tag{1}$$

In this equation:

- $V^{\pi}(s)$ represents the value of state s under policy π.
- $\pi(a|s)$ is the probability of taking action a in state s under policy π.
- $p(s', r|s, a)$ is the transition probability from state s to state s' with reward r after taking action a.
- γ is the discount factor.
- $V^{\pi}(s')$ is the value of the next state s' under policy π.

This equation describes how the value of a state under a policy is the sum of the expected immediate reward and the discounted value of the next state, weighted by the transition probabilities and the policy's action probabilities.