P3: Exploration Strategies

Plot error to see converge?

VI: Epsilon, Gamma

PI: Epsilon, Gamma

Q: gamma, alpha, alpha\_decay, alpha\_min, epsilon, epsilon\_min, epsilon\_decay, n\_iter

How does discount affect convergence?

I can see here higher discount takes more iterations to convergence ?

How is that so?

we should do this (vary the size of each MDP) because a grid world and non-grid world problem might scale differently in terms of how policy and value iteration handle it.

probs\_f, rewards\_f = hiive.mdptoolbox.example.forest(S=3000, p=0.001)