Q-Learning

- Q-Learning: sample-based Q-value iteration

$$Q_{k+1}(s,a) \leftarrow \sum_{s'} T(s,a,s')\left[R(s,a,s') + \gamma \max_{a'} Q_k(s',a')\right]$$

- Learn Q(s,a) values as you go
  - Receive a sample (s,a,s’,r)
  - Consider your old estimate: $Q(s,a)$
  - Consider your new sample estimate:

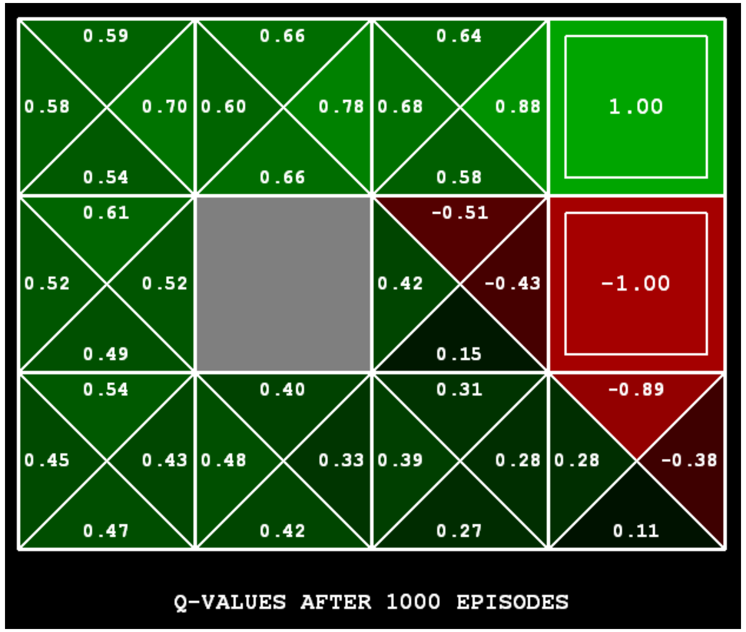
$$sample = \underbrace{R(s,a,s')}_{\text{reward}} + \gamma \max_{a'} Q(\overset{\text{next State}}{s'}, a')$$

  - Incorporate the new estimate into a running average:

Your **update** Function will implement this

$$\underbrace{Q(s,a)}_{\text{updated value}} \leftarrow (1-\alpha)\underbrace{Q(s,a)}_{\text{old value}} + (\underset{\text{learning rate}}{\alpha})\,[sample]$$

$$Q(s,a) \leftarrow Q(s,a) + \alpha(\,sample - Q(s,a)\,)$$



Q-VALUES AFTER 1000 EPISODES

[Demo: Q-learning – gridworld (L10D2)]
[Demo: Q-learning – crawler (L10D3)]