When it got reward r and the situation changed to s' after it took an action a in situation s, it revises the Q value for s and a as

$$\Delta \mathrm{Q}(s,a) = \alpha \cdot \left(r + \gamma \max_{b \in \mathrm{A}(s')} \mathrm{Q}(s',b) - \mathrm{Q}(s,a) \right)$$

under the learning rate α ($0 < \alpha < 1$) and discount rate γ ($0 < \gamma < 1$).