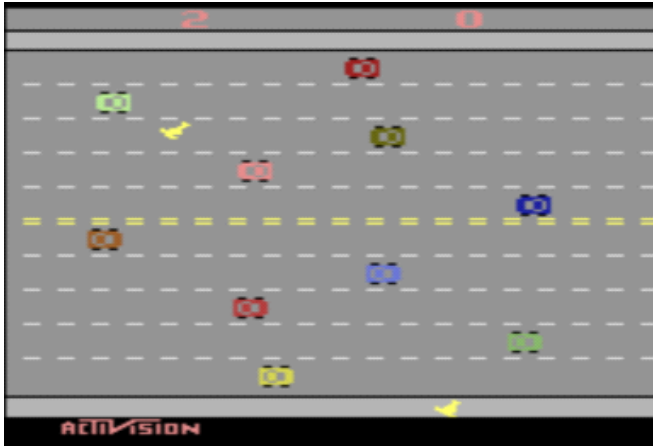


Reward Shaping

Friday, 21 September 2018 15:22 PM



shaped reward

Key idea: add some small additional reward for particular behaviour:

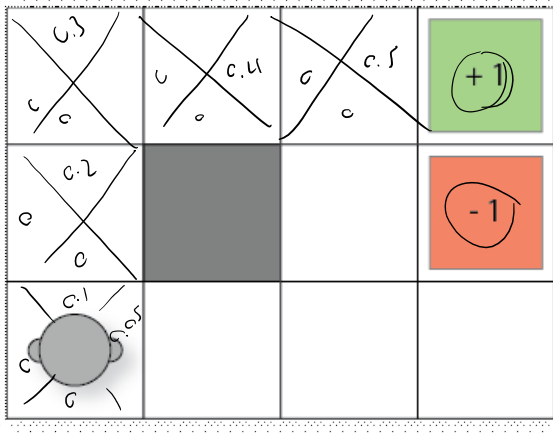
$$Q(s,a) := Q(s,a) + \alpha [r + \gamma \max_{a'} Q(s',a') - Q(s,a)]$$

$$0.1 \times 10 - 0.1 \times 0$$

$$F(s,s') = \gamma P(s') - P(s)$$

GridWorld example:

$$P(s) = 1 / (|x(g) - x(s)| + |y(g) - y(s)|)$$



$$F((1,2), (2,2)) = 0.9 \times \frac{1}{1} - \frac{1}{2} = 0.4$$

$$F((0,0), (0,1)) = 0.9 \times 1/4 - 1/5 = 0.025 \text{ -- in this case the reward is misleading!}$$

Alternative to reward shaping: Q-function initialisation