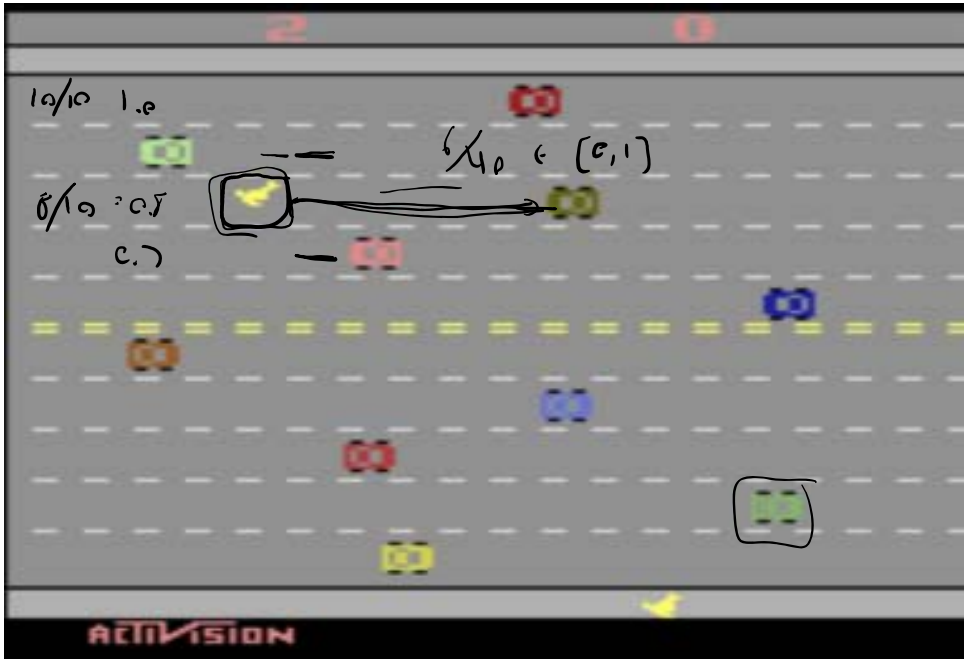


Q-Function Approximation



$$12 \times 40 = 480$$

$$480^2 + 2^{480} =$$

$$3.12 \times 10^{144}$$

States

$\times |A|$

Features and representation:

Feature function returning a feature vector:

$$\rightarrow \underline{f(s,a)} = \begin{pmatrix} f_1(s,a) \\ f_2(s,a) \\ \vdots \\ f_n(s,a) \end{pmatrix}$$

Weight vector:

$$n \times |A|$$

$$f(s) = \begin{pmatrix} f_1(s) \\ \vdots \\ f_n(s) \end{pmatrix}$$

$$f(s,a) = \begin{pmatrix} f_{1a}(s,a) \\ f_{2a}(s,a) \\ \vdots \\ f_{na}(s,a) \end{pmatrix}$$

Approximating $Q(s,a)$ with linear regression

$$Q(\underline{s}, \underline{a}) = \underline{w}_1 \cdot \underline{f}_1(\underline{s}, \underline{a}) + \underline{w}_2 \cdot \underline{f}_2(\underline{s}, \underline{a}) + \dots$$

$$\underline{Q(s,a)} = \underline{0.31} \cdot \underline{f_1(s,a)} + \underline{0.17} \cdot \underline{f_2(s,a)} + \dots + \underline{0.04} \cdot \underline{f_n(s,a)}$$

$$= \underline{0.9} \quad \underline{0.9} \quad \underline{0.9}$$

Q-learning Update:

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

$$s \xrightarrow{a} s' \quad \text{for each feature } i$$

$$w_i^a = w_i^a + \alpha[r + \gamma \cdot \underline{Q(s',a')} - Q(s,a)] f_i(s,a)$$

$$0.9$$

$$w_i^a = w_i + \alpha [r + \gamma \cdot \underbrace{Q(s', a')}_{0.9}] f_i(s, a)$$

Example (Freeway):

Assume $Q(s, a) = 0$ for all s, a

$$\begin{aligned} \underline{w_{14}^{u_1}} &= 0 + 0.5 [10 + 0.9 \times 0] \cdot 0.9 \\ &= 0.5 \cdot 0.9 \\ &= 0.45 \end{aligned}$$

$$\begin{aligned} Q(s, u_p) &= 0.45 \cdot \underline{f_i(s, u_p)} + \dots \\ &= 0.45 \cdot 0.1 + \dots \cdot 0.1 \\ \bigcirc &= 0.05 \end{aligned}$$

Update:

$$w_i \leftarrow w_i + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] f_i(s, a)$$

$$\vartheta_i \leftarrow \vartheta_i + \alpha [r + \gamma \max_{a'} Q(s', a') - Q(s, a)] \partial Q(s, a) / \partial \vartheta \quad , s)$$

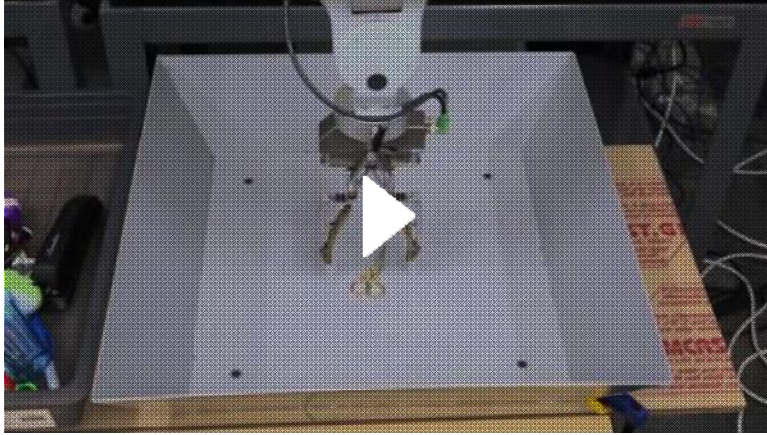
, s)

Deep Q learning:

$$\begin{aligned} &\vartheta \quad Q(\bar{s}, \bar{a}; \vartheta) \\ \vartheta &= \vartheta + \alpha [r + \gamma \cdot \max_{\bar{a}'} Q(\bar{s}', \bar{a}') - Q(\bar{s}, \bar{a})] \nabla Q(\bar{s}, \bar{a}; \vartheta) \end{aligned}$$

Deep Q-learning for robot grasping:

[Learning Hand-Eye Coordination for Robotic Grasping](#)



Strengths and Limitations

Strengths

- compact representation
- Q-value propagation

Limitations

- approximation

