$$2n \begin{cases} n \\ n \end{cases}$$

$$1 \qquad 2 \qquad R^{n\times 2}$$

$$\vec{z} = n+1$$

$$(1,2)$$

$$\underline{(a,b)}$$

$$\vec{x} \longrightarrow \vec{x}\,' \begin{bmatrix} | & | & | \end{bmatrix}$$

$$P_{\pi\cdots} = \sum_{a \in A} \pi_\eta(a)\, P_{ahh'}$$

TD Loss

$$L(\theta, \theta^-; D) = \sum_{(\phi, a, r, \phi') \in D} \left(\overset{\text{reward}}{r} + \max_{a' \in A} \hat{Q}_{\theta^-}(\phi', a') - \hat{Q}_{\theta}(\phi, a) \right)$$

temporal
difference

$(h,a)\big)^2$