

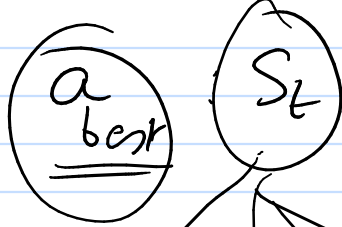
$$\textcircled{1} \quad Q(s, a) = Q(s, a) + \alpha \left(\underbrace{r + \gamma \max_a Q(s', a)}_{\text{TD error}} - Q(s, a) \right)$$

$$\textcircled{2} \quad \text{batch} \quad \underline{w}^{t+1} = \underline{w}^t - \alpha \left(\underbrace{\text{TD error}}_{\text{TD error}} \right) \nabla_w Q(s, a; w)$$

$$\frac{d}{dw} \left(r + \gamma \max_a Q(s', a; \underline{w}) - Q(s, a; w) \right)^2$$

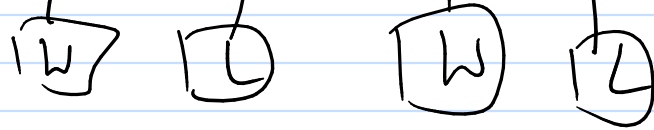
Diagram illustrating the derivative of the squared TD error with respect to the weight vector w . The expression is $\frac{d}{dw} \left(r + \gamma \max_a Q(s', a; \underline{w}) - Q(s, a; w) \right)^2$. The term \underline{w} is circled, and an arrow points from it to the w in the second term of the squared expression. Another arrow points from the circled \underline{w} to the w in the gradient term of the equation above.

Goal:



\checkmark TT(S)

MCS



$\frac{2}{4}$

