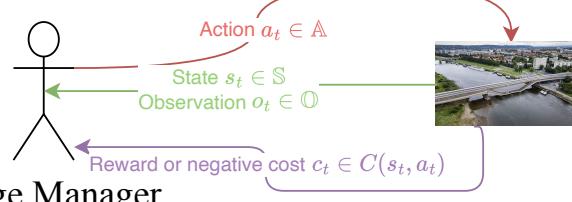
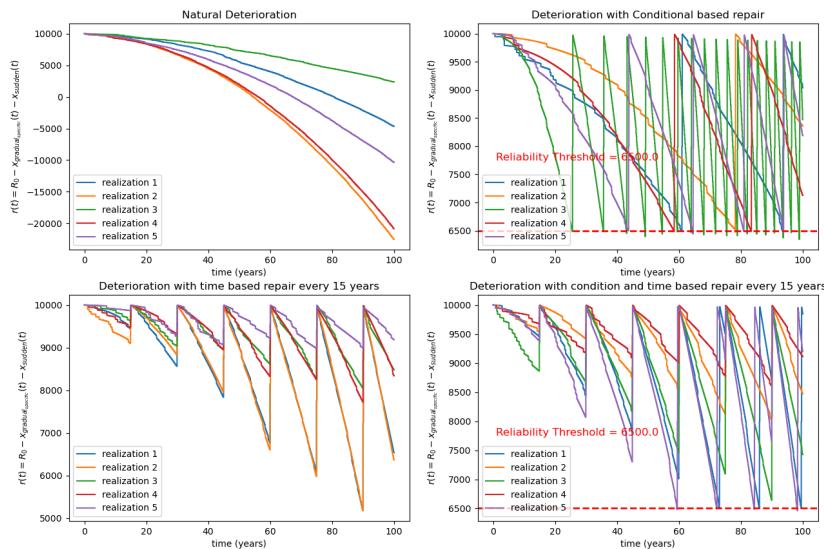


Mathematic Modclng: Partially Observable Markov Decision Process (POMDP) $\mathbb{S}, \mathbb{O}, \mathbb{A}, \mathbf{T}, \mathbf{O}, \mathbf{R}, \gamma$



The key is to find the policy which could minimize the expected total cost $\pi^* = \arg \max_{\pi} Q^{\pi}(s_t, a_t)$



Deep Neural Network

