Topics in machine learning

Siddharth Bhat

Monsoon 2019

Contents

Chapter 1

Policy iteration

$$\pi_{k+1}(s) = arg \max_{a \in A(s)} r(s,a) + \gamma \sum_{s} P(s'|s,a) v_{\pi_k}(s')$$

Theorem 1 *The policy iteration algorithm generates a sequecnce of policies with non-decreasing state values. That is, $V^{\pi_{k+1}} \geq V^{\pi_k}$, $V^{\pi} \in \mathbb{R}^n$, is the vector of state values for state π*

Proof 1 *F^{π_k} is the bellman expectation operator (?)*
Since V^{π_k} is a fixed point of F^{π_k},

$$\begin{aligned}
V^{\pi_k} = F^{\pi_k}(V^{\pi_k}) \leq F(V^{\pi_k}) \qquad & \textit{(upper bounded by max value)} \\
F(V^{\pi_k}) = F^{\pi_{k+1}}(V^{\pi_k}) \qquad & \textit{(By defn of policy improvement step)} \\
V^{\pi_k} \leq F^{\pi_{k+1}}(V^{\pi_k}) \qquad & \textit{(eqn 1)} \\
F^{\pi_{k+1}}(V^{\pi_k}) \leq (F^{\pi_{k+1}})^2(V^{\pi_k}) \qquad & \textit{Monotonicity of } F^{\pi_{k+1}} \\
\forall t \geq 1,\ F^{\pi_{k+1}}(V^{\pi_k}) \leq (F^{\pi_{k+1}})^t(V^{\pi_k}) \qquad & \textit{(Monotonicity of } F^{\pi_{k+1}}\textit{)} \\
F^{\pi_{k+1}}(V^{\pi_k}) \leq (F^{\pi_{k+1}})^t(V^{\pi_k}) \leq V^{\pi_{k+1}} \qquad & \textit{(Contraction mapping, } V^{\pi_{k+1}} \textit{ is fixed point)}
\end{aligned}$$