Review of Continuous Adaptation Via Meta-Learning in Nonstationary and Competitive Environments

Guannan Hu

April 17, 2018

1 A Probabilistic View of Model-Agnostic Meta-LEarning (MAML)

Assume that we are given a distribution over tasks, $\mathcal{D}(T)$, where each task, T, is a tuple:

$$T := (L_T, P_T(\mathbf{x}), P_T(\mathbf{x}_{t+1}|\mathbf{x}, \mathbf{a}_t), H) \tag{1}$$

L_T is a task specific loss function that maps a trajectory, $\boldsymbol{\tau} := (\mathbf{x}_0, \mathbf{a}_1, \mathbf{x}_1, R_1, ..., \mathbf{a}_H, \mathbf{x}_H, R_H) \in \mathcal{T}$, to a loss value, i.e., $L_T : \mathcal{T} \to \mathbb{R}$; $P_T\mathbf{x}$ and $P_T(\mathbf{x}_{t+1}|\mathbf{x}_t, \mathbf{a}_t)$ define the Markovian dynamics of the environment in task T; H denotes the horizon; observations, $\mathbf{x}_t$, and actions, $\mathbf{a}_t$, are elements (typically, vectors) of the observation space, $\mathcal{X}$, and action space, $\mathcal{A}$, respectively. The loss of a trajectory, $\boldsymbol{\tau}$, is the negative cumulative reward, $L_T(\boldsymbol{\tau}) := -\sum_{t=1}^{H} R_t$.