Overestimation

Yanhua Huang

Jul 2019

Maximum operation over noisy estimation causes overstimation, which can be addressed by Double Q Estimation for Q-Learning and TD3 (Twin Delayed Deep Deterministic Policy Gradients) for Actor-Critic in reinforcement learning.

Similar problems is learned in statistic. The definition of variance is $\sigma^2 = \mathbb{E}[(x-\mu)^2]$. With Monte Carlo method, its estimation is

$$\hat{\sigma}^2 = \frac{1}{N}\sum_{i=1} N(x_i - \mu)^2. \tag{1}$$

However, it is underestimated. A more accurate estimation is $\frac{N}{N-1}\hat{\sigma}^2$, which is used in the inference part of Batch Normalization. This derivative process is simple. Consider

$$\mathbb{E}[\hat{\sigma}^2] = \mathbb{E}[x_i^2 - 2x_i\mu + \mu^2], \tag{2}$$

the samples are often viewed as idd. For $i \neq j$, $\mathbb{E}[x_i x_j] = \mathbb{E}[x_i]\mathbb{E}[x_j] = \mu^2$. With the definition of variance, $\mathbb{E}[x_i^2] = \mu^2 + \sigma^2$,

$$N\mathbb{E}[\hat{\sigma}^2] = (N-1)\sigma^2. \tag{3}$$