RL: Lecture 21—Approximation

Harvey Mudd College

April 15, 2020

Neil Rhodes

Overview

- Why approximation?
- Value-function Approximation
- Prediction Objective
- Linear methods
- Stochastic Gradient Descent (SGD) methods

Why approximation?

Value-function Approximation

weight vector $\boldsymbol{w} \in \mathbb{R}^d$

approximate value function: $\hat{v}(\boldsymbol{s}, \boldsymbol{w}) \approx v_\pi(\boldsymbol{s})$

$$\boldsymbol{s} \mapsto \boldsymbol{u}$$

$\boldsymbol{s}$: state being updated

$\boldsymbol{u}$: update target $\boldsymbol{s}$'s value is being shifted toward

Prediction Objective

How much we care about the error at each state: state distribution $\mu(s) \geq 0, \sum_s \mu(s) = 1$

Mean Squared Value Error: $\overline{VE}(\boldsymbol{w}) = \sum_{s \in S} \mu(s)[v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2$

On-policy distribution in episodic tasks

$\mu(s)$: fraction of the time spent in s under policy π.

Probability starting in state s: $h(s)$

time spent in s: $\eta(s) = h(s) + \sum_{\bar{s}} \eta(\bar{s}) \sum_{a} \pi(a|\bar{s}) p(s|\bar{s}, a)$

Normalized to sum to one: $\mu(s) = \frac{\eta(s)}{\sum_{s'} \eta(s')}$

Goal for $\overline{VE}$

Find a global optimum:
$\boldsymbol{w}^*$ such that $\forall \boldsymbol{w} : \overline{VE}(\boldsymbol{w}^*) \leq \overline{VE}(\boldsymbol{w})$

May have to settle for local optimium:
$\boldsymbol{w}^*$ such that $\forall \boldsymbol{w}$ **in neighborhood of** $\boldsymbol{w}^* : \overline{VE}(\boldsymbol{w}^*) \leq \overline{VE}(\boldsymbol{w})$

Gradient

Gradient: operation that takes a function of a vector and creates a vector of partial derivatives:

$$\nabla f(\boldsymbol{w}) = \left[\frac{\partial f(\boldsymbol{w})}{\partial w_1}, \frac{\partial f(\boldsymbol{w})}{\partial w_2}, \cdots, \frac{\partial f(\boldsymbol{w})}{\partial w_d}\right]^T$$

SGD

Given loss function $\overline{VE}(\boldsymbol{w})$ which consists of a sum of individual losses: $\sum_{s \in S} \mu(s)[v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2$

1. Pick an $\boldsymbol{s} \in S$ according to $\boldsymbol{\mu}$
2. Find loss for that $\boldsymbol{s}$: $[v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2$
3. Find gradient of individual loss w.r.t. $\boldsymbol{w}$:
   $\nabla[v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2 = 2(v_\pi(s) - \hat{v}(s, \boldsymbol{w}))\nabla\hat{v}(s, \boldsymbol{w})$
4. Reduce individual loss by adjusting $\boldsymbol{w}$ in opposite direction of gradient:
   $\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \frac{1}{2}\alpha\nabla[v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2 =$
   $\boldsymbol{w}_t + \alpha(\hat{v}(s, \boldsymbol{w}) - v_\pi(s))\nabla\hat{v}(s, \boldsymbol{w})$

Linear Methods

$\boldsymbol{x}(s)$: feature vector of $\boldsymbol{s}$

$$\hat{v}(\boldsymbol{w}) = \boldsymbol{w}^T \boldsymbol{x}(s) = \sum_{i=1}^{d} w_i x_i(s)$$

SGD—Linear methods

For any weight, w_i, $\dfrac{\partial \hat{v}}{\partial w_i} = x_i$

Gradient of $\hat{v}$ with respect to $\boldsymbol{w}$:

$$\nabla \hat{v}(s, \boldsymbol{w}) = \left[\frac{\partial \hat{v}(s, \boldsymbol{w})}{\partial w_1}, \frac{\partial \hat{v}(s, \boldsymbol{w})}{\partial w_2}, \cdots, \frac{\partial \hat{v}(s, \boldsymbol{w})}{\partial w_d} \right] = \boldsymbol{x}(s)^T$$

To reduce $\overline{VE}(\boldsymbol{w})$, tweak $\boldsymbol{w}$ in a direction such that $\overline{VE}(\boldsymbol{w})$ is reduced

SGD—Linear methods

Given loss function $\overline{VE}(\boldsymbol{w}))$ which consists of a sum of individual losses: $\sum_{s \in S} \mu(s)[v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2$

1. Pick an $\boldsymbol{s} \in S$ according to μ
2. Find loss for that $\boldsymbol{s}$: $[v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2$
3. Find gradient of individual loss w.r.t. $\boldsymbol{w}$:
   $\nabla[v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2 = 2(v_\pi(s) - \hat{v}(s, \boldsymbol{w}))\nabla\hat{v}(s, \boldsymbol{w})$
4. Reduce individual loss by adjusting $\boldsymbol{w}$ in opposite direction of gradient:
   $\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \frac{1}{2}\alpha\nabla[v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2 =$
   $\boldsymbol{w}_t + \alpha(\hat{v}(s, \boldsymbol{w}) - v_\pi(s))\nabla\hat{v}(s, \boldsymbol{w}) = \boldsymbol{w}_t + \alpha(\hat{v}(s, \boldsymbol{w}) - v_\pi(s))\boldsymbol{x}(s)$