Overview

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

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

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

$s \mapsto u$

$s$: state being updated

$u$: update target $s$'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: \( \overline{VE}(\mathbf{w}) = \sum_{s \in S} \mu(s)[v_\pi(s) - \hat{v}(s, \mathbf{w})]^2 \)
On-policy distribution in episodic tasks

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

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 optimum:
$\boldsymbol{w}^*$ such that $\forall \boldsymbol{w} \text{ in neighborhood of } \boldsymbol{w}^* : \overline{VE}(\boldsymbol{w}^*) \leq \overline{VE}(\boldsymbol{w})$
Gradient

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

\[ \nabla f(w) = \begin{bmatrix} \frac{\partial f(w)}{\partial w_1}, & \frac{\partial f(w)}{\partial w_2}, & \cdots, & \frac{\partial f(w)}{\partial w_d} \end{bmatrix}^T \]
SGD

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

1. Pick an $s \in S$ according to $\mu$

2. Find loss for that $s$: $[v_\pi(s) - \hat{v}(s, \mathbf{w})]^2$

3. Find gradient of individual loss w.r.t. $\mathbf{w}$: $\nabla[v_\pi(s) - \hat{v}(s, \mathbf{w})]^2 = 2(v_\pi(s) - \hat{v}(s, \mathbf{w})) \nabla \hat{v}(s, \mathbf{w})$

4. Reduce individual loss by adjusting $\mathbf{w}$ in opposite direction of gradient:

$$\mathbf{w}_{t+1} = \mathbf{w}_t - \frac{1}{2} \alpha \nabla[v_\pi(s) - \hat{v}(s, \mathbf{w})]^2 = \mathbf{w}_t + \alpha(\hat{v}(s, \mathbf{w}) - v_\pi(s)) \nabla \hat{v}(s, \mathbf{w})$$
Linear Methods

\[ x(s) : \text{feature vector of } s \]

\[ \hat{v}(w) = w^T x(s) = \sum_{i=1}^{d} w_i x_i(s) \]
SGD—Linear methods

For any weight, \( w_i \), \( \frac{\partial \hat{v}}{\partial w_i} = x_i \)

Gradient of \( \hat{v} \) with respect to \( w \):
\[
\nabla \hat{v}(s, w) = \left[ \frac{\partial \hat{v}(s, w)}{\partial w_1}, \frac{\partial \hat{v}(s, w)}{\partial w_2}, \ldots, \frac{\partial \hat{v}(s, w)}{\partial w_d} \right] = \mathbf{x}(s)^T
\]

To reduce \( \overline{VE}(w) \), tweak \( w \) in a direction such that \( \overline{VE}(w) \) is reduced.
SGD—Linear methods

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

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