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?

tabular

must enumerate all either state/value

or

state/action/value combination

Memory issue storing table

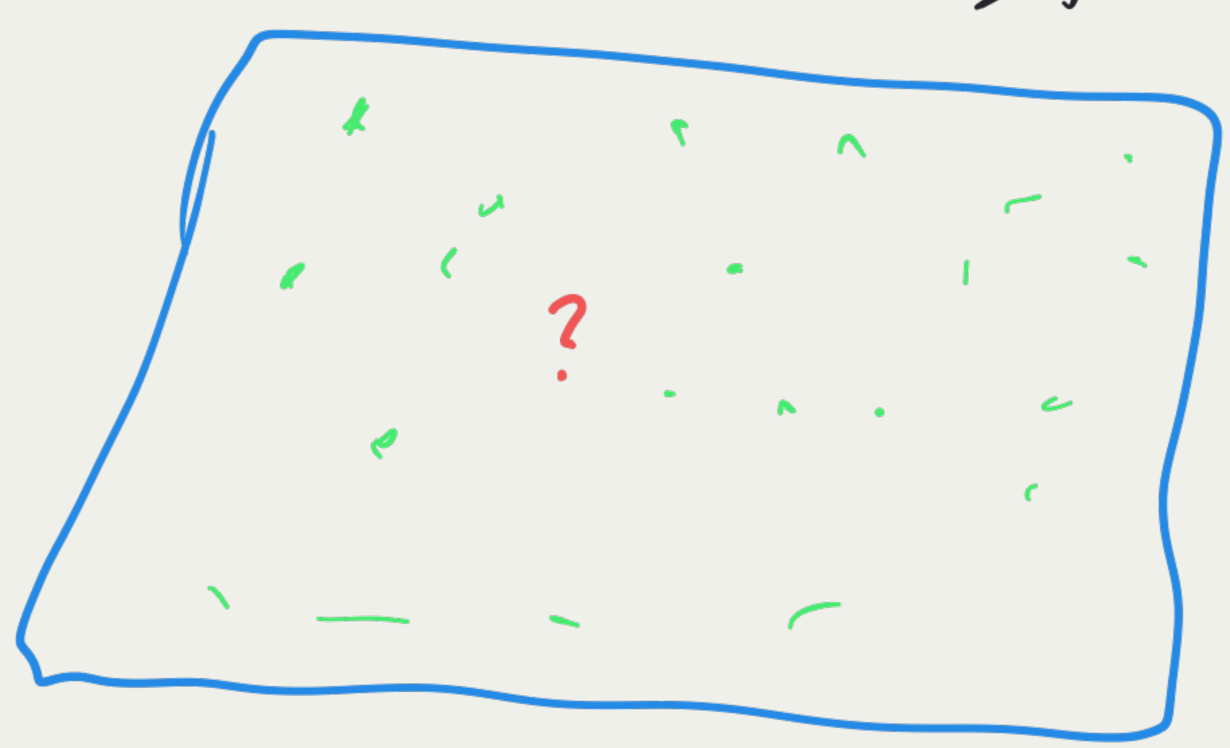
CPU issue filling in the table

self-driving cars

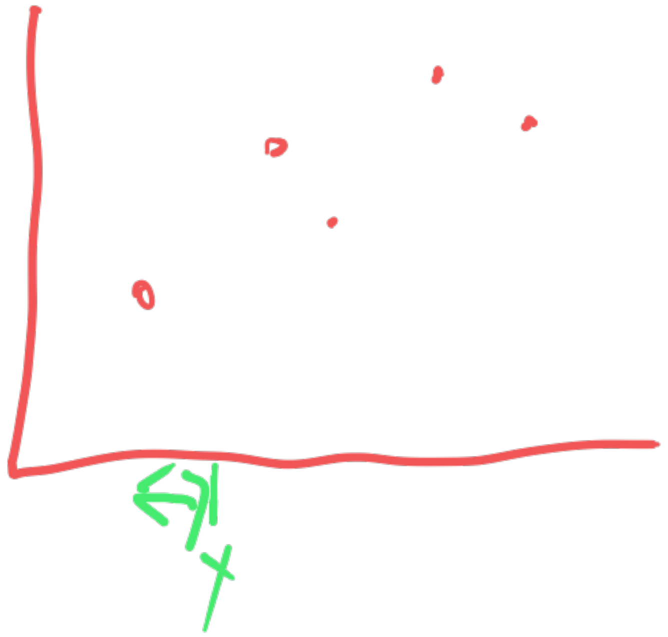
rotation of steering wheel

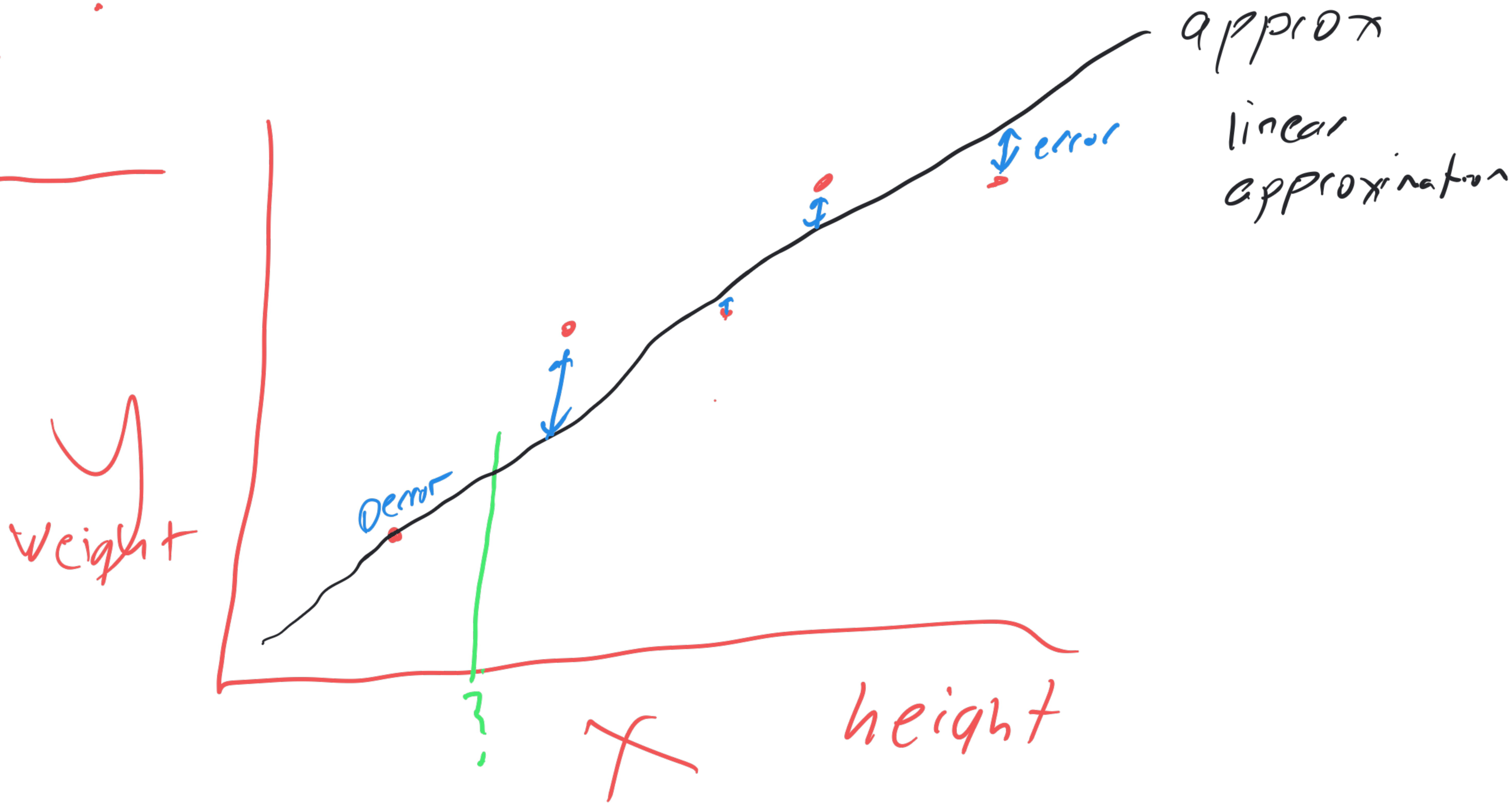
$\mathbb{R} \rightarrow$ inf possibilities

chess lots of different states



huge





linear approx saving space

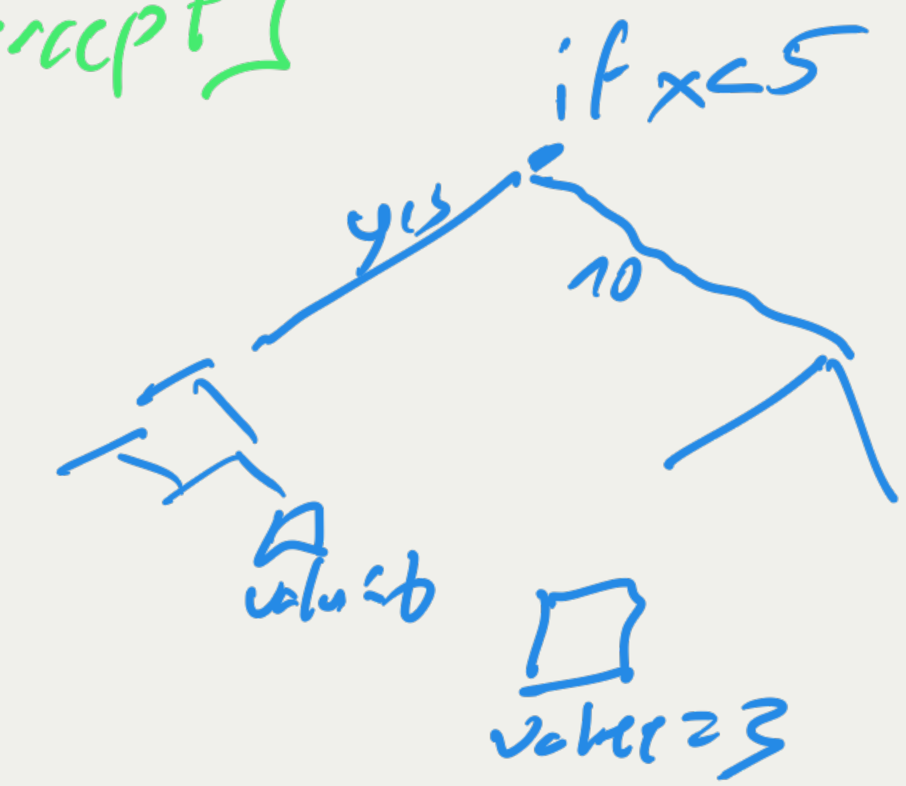
$\underbrace{\text{save slope \& intercept}}_{\text{params/weights}}$

tabular

	x	y
1		
2		
3		
4		
5		

Value-function Approximation

linear case $w = [\text{slope} \;\; \text{intercept}]$



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

approximate value function: $\hat{v}(\underbrace{\boldsymbol{s}}_{\text{state}}, \underbrace{\boxed{\boldsymbol{w}}}_{\text{learned weights}}) \approx v_\pi(\boldsymbol{s})$

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

$\boldsymbol{s}$: state being updated

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

$$v(s) = v(s) + \alpha[R + v(s') - v(s)]$$

$$v(s) = v(s) + \alpha[\underbrace{\text{target}}_{\text{target}} - \underbrace{v(s)}_{\text{current value}}]$$

Calculating new target
(may use $\hat{v}$ rather than v_π)

update $\hat{v}(s, w)$

↓ from

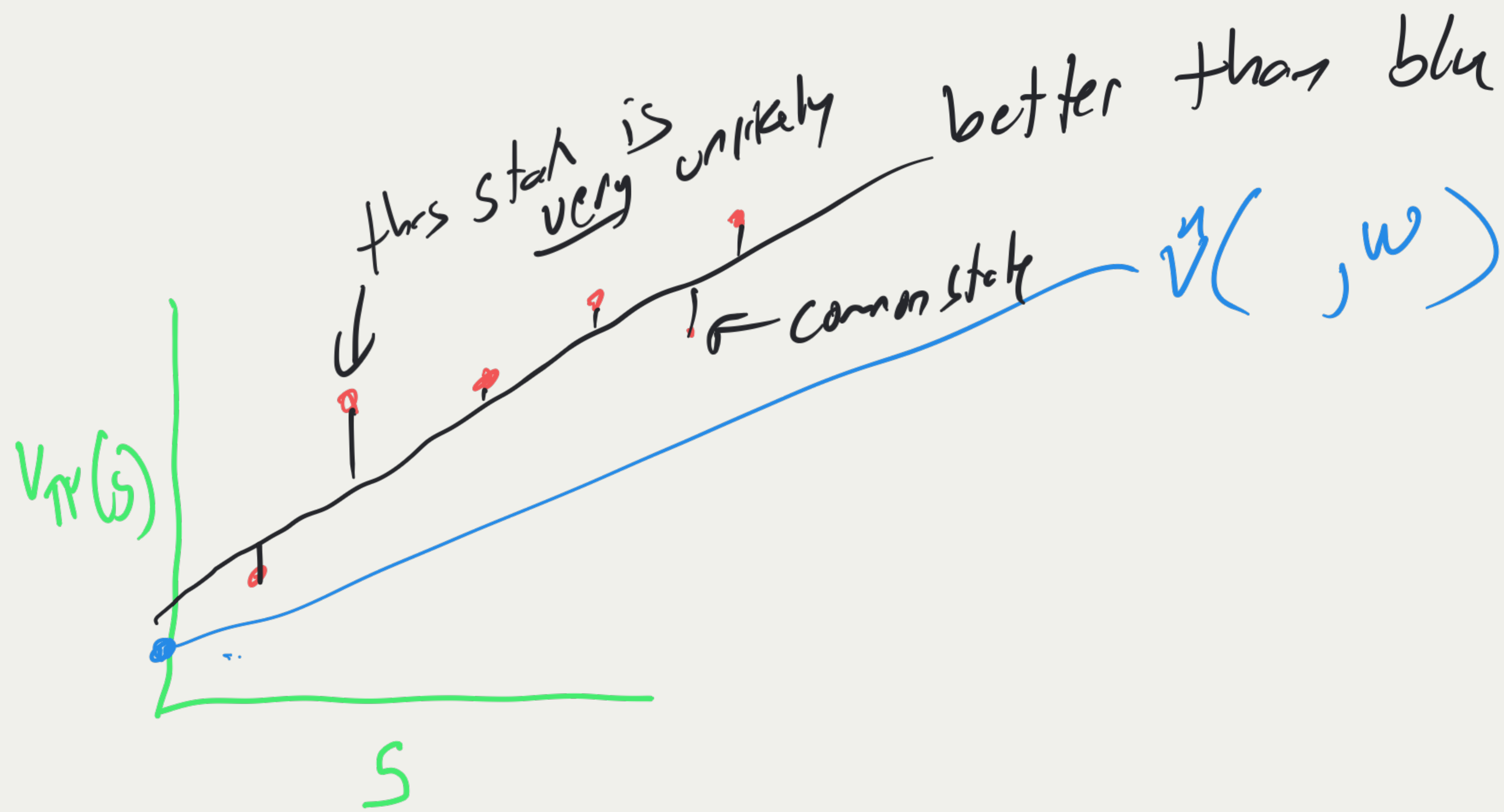
$\hat{v}(s, w)$ more towards u

updating weight vector over time

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: $h(s)$

prob of being in
predecessor state

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

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

Goal for $\overline{VE}$

value error function

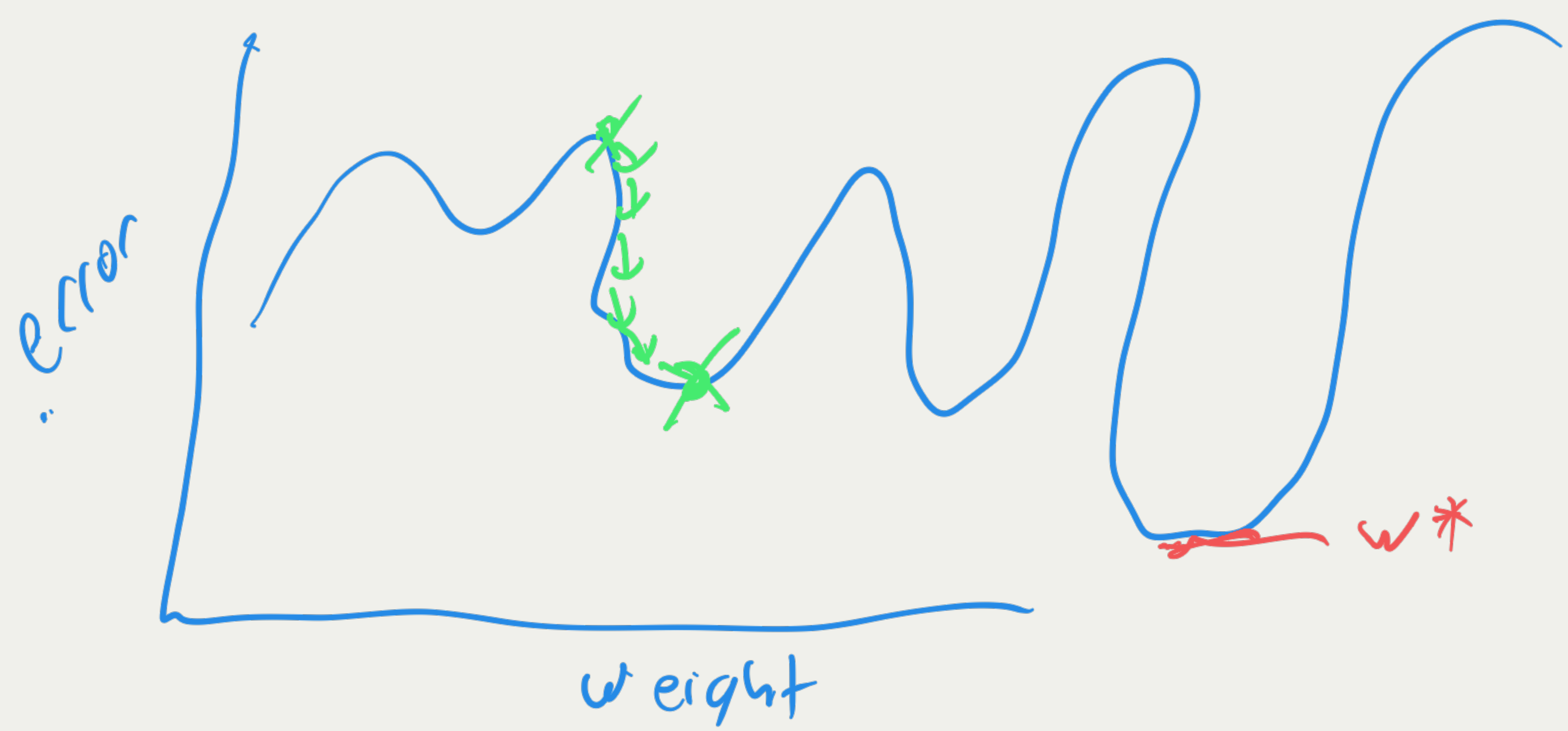
(loss function)

Find a global optimum:

$\boxed{\boldsymbol{w}^*}$ 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 $\boldsymbol{w}^*$: $\overline{VE}(\boldsymbol{w}^*) \leq \overline{VE}(\boldsymbol{w})$



$$f(x,y) = 2xy$$

$$\frac{\partial f}{\partial x} = 2y \qquad \frac{\partial f}{\partial y} = 2x$$

$$f\left(\begin{bmatrix} 2 \\ 3 \end{bmatrix}\right) = 2 \cdot 2 \cdot 3 = 12$$

Gradient

$$\boxed{f(w) = 2w_1 w_2}$$

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

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

nabla or del

$$\nabla f(w) = \left[\frac{\partial f(w)}{\partial w_1} \quad \frac{\partial f(w)}{\partial w_2}\right]$$

$$= \left[2w_2 \quad 2w_1\right]$$

SGD

Stochastic Gradient Descent → just sample one state at a time
vs.
Gradient Descent: we have a set S of states & their associated $v_\pi(s)$

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

equiprobable → $\overline{VE}(w) = \frac{1}{|S|} \sum_{s \in S} [v_\pi(s) - \hat{v}(s,w)]^2$

$\nabla \overline{VE}(w)$ measures what?
[for each component of w]
how to increase loss

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

   reduce loss

   baby steps

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

   (approximate − actual)

Avg M

$$\overline{VE}(w) = \frac{1}{|S|}$$

Avg MSE

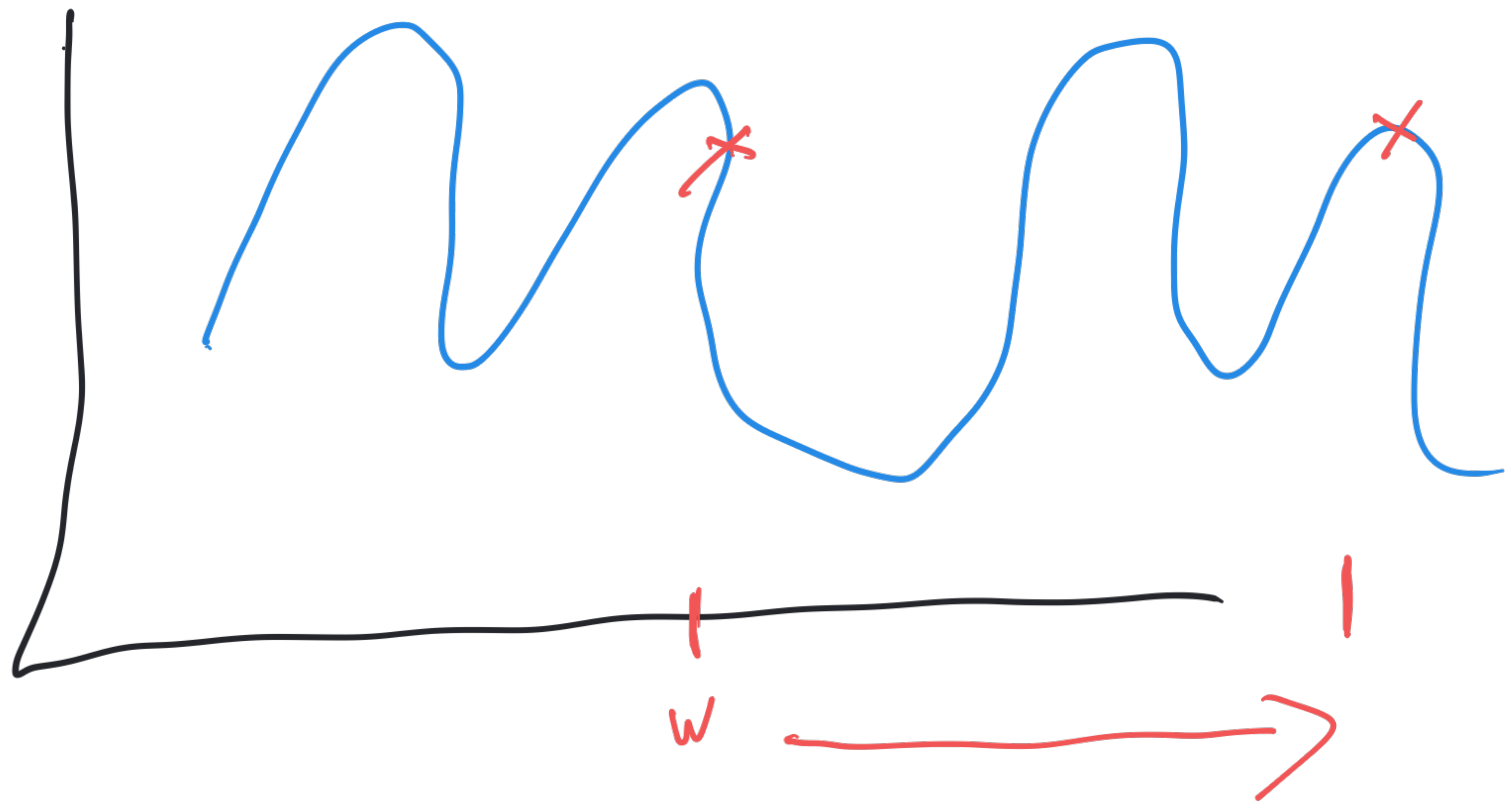
$$\overline{VE}(w) = \frac{1}{|S|}\sum_{s\in S}\left[v_\pi(s) - \hat{v}(s,w)\right]^2$$

$$\frac{\partial \overline{VE}(w)}{\partial w_i} = \frac{1}{|S|}\sum_{s\in S}\frac{\partial \left[v_\pi(s) - \hat{v}(s,w)\right]^2}{\partial w_i}$$

$$= \frac{1}{|S|}\sum_{s\in S} 2\left[v_\pi(s) - \hat{v}(s,w)\right]\cdot -\frac{\partial \hat{v}(s,w)}{\partial w_i}$$

$$= \frac{1}{|S|}\sum_{s\in S} 2\left(\hat{v}(s,w) - v_\pi(s)\right)\frac{\partial \hat{v}(s,w)}{\partial w_i}$$

$$\nabla \overline{VE}(w) = \frac{1}{|S|}\boxed{\sum_{s\in S} 2\left(\hat{v}(s,w) - v_\pi(s)\right)}\nabla \hat{v}(s,w)$$



w

N/S E/W

location $\{x, y\}$

height (location)

$$\nabla \text{height} = \left[\frac{\partial\, loc}{\partial x} \quad \frac{\partial\, loc}{\partial y}\right]$$

particular loc
our gradient is

$\{5, -3\}$

ball moving down in x axis
up in y-axis

$x+ \rightarrow \overrightarrow{pos +}$
$y+ \rightarrow pos -$

Linear Methods

s = a chess board

$x_1(s)$ = # pawns white

$x_2(s)$ = # knights white

⋮

$x_{11}(s)$ = in check white

$x_{13}(s)$ = mobility in middle of board

extract from boardconfig

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

d elements

$$\begin{bmatrix} 0 \\ 5 \\ 1 \\ 3 \\ \vdots \\ 8 \end{bmatrix}$$

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

$\hat{v}(w, s)$

$x = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix}$ $w = \begin{bmatrix} 1 \\ 2 \\ -1 \end{bmatrix}$

$\hat{v}(w, s) = \begin{bmatrix} 1 & 2 & -1 \end{bmatrix} \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} = 3$

SGD—Linear methods

$\hat{v} = \sum_{i=1}^{d} w_i x_i$

$w_1 x_1 + w_2 x_2 + \cdots w_d x_d$

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

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

$$\nabla \hat{v}(\boldsymbol{s}, \boldsymbol{w}) = \left[\underbrace{\frac{\partial \hat{v}(\boldsymbol{s}, \boldsymbol{w})}{\partial w_1}}_{x_1(s)}, \underbrace{\frac{\partial \hat{v}(\boldsymbol{s}, \boldsymbol{w})}{\partial w_2}}_{x_2(s)}, \cdots, \underbrace{\frac{\partial \hat{v}(\boldsymbol{s}, \boldsymbol{w})}{\partial w_d}}_{x_d(s)} \right] = \boldsymbol{x}(\boldsymbol{s})^T$$

$= x(s)^T$

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 $\boldsymbol{\mu}$
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:
   $$\boldsymbol{w}_{t+1} = \boldsymbol{w}_t - \frac{1}{2}\alpha\nabla[v_\pi(s) - \hat{v}(s, \boldsymbol{w})]^2 =$$
   $$\boldsymbol{w}_t + \alpha(\hat{v}(s, \boldsymbol{w}) - v_\pi(s))\boxed{\nabla\hat{v}(s, \boldsymbol{w})} = \boldsymbol{w}_t + \alpha(\hat{v}(s, \boldsymbol{w}) - v_\pi(s))\boxed{\boldsymbol{x}(s)}$$