Prof. Dr. Leif Döring
André Ferdinand, Sara Klein

Reinforcement Learning

3. Excercise Sheet

1. Best Baseline

The variance of a random vector X is defined by to be $\mathbb{V}[X] := \mathbb{E}[||X||_2^2] - ||E[X]||_2^2$. Show by differentiation that

$$b_* = \frac{\mathbb{E}_{\pi_\theta}[X_A||\nabla \log \pi_\theta(A)||_2^2]}{\mathbb{E}_{\pi_\theta}[||\nabla \log \pi_\theta(A)||_2^2]}$$

is the baseline that minimises the variance of the unbiased estimators

$$(X_A - b)\nabla \log(\pi_\theta(A)), \quad A \sim \pi_\theta,$$

of $\nabla J(\theta)$.

Solution:

We have

$$\begin{aligned}
&\mathbb{V}\Big((X_A - b)\nabla \log(\pi_\theta(A))\Big)\\
&= \mathbb{E}\Big[(X_A - b)^2||\nabla \log(\pi_\theta(A))||_2^2\Big] - \Big\|\mathbb{E}\Big[(X_A - b)\nabla \log(\pi_\theta(A))\Big]\Big\|_2^2\\
&= \mathbb{E}\Big[(X_A - b)^2||\nabla \log(\pi_\theta(A))||_2^2\Big] - \Big\|\mathbb{E}\Big[X_A\nabla \log(\pi_\theta(A))\Big]\Big\|_2^2,
\end{aligned}$$

where we used the baseline trick in the last equation. We define $f(A) = ||\nabla \log(\pi_\theta(A))||_2$ *to have a better overview. Then*

$$\begin{aligned}
&\mathbb{V}\Big((X_A - b)\nabla \log(\pi_\theta(A))\Big)\\
&= \mathbb{E}\Big[(X_A - b)^2 f(A)^2\Big] - \Big\|\mathbb{E}\Big[X_A f(A)\Big]\Big\|_2^2\\
&= \mathbb{E}\Big[X_A^2 f(A)^2\Big] - 2b\mathbb{E}\Big[X_A f(A)^2\Big] + b^2\mathbb{E}\Big[f(A)^2\Big] - \Big\|\mathbb{E}\Big[X_A f(A)\Big]\Big\|_2^2
\end{aligned}$$

We calculate the first derivative

$$\begin{aligned}
&\frac{\partial \mathbb{V}\Big((X_A - b)\nabla \log(\pi_\theta(A))\Big)}{\partial b}\\
&= -2\mathbb{E}\Big[X_A f(A)^2\Big] + 2b\mathbb{E}\Big[f(A)^2\Big].
\end{aligned}$$

Solving for the root gives

$$b* = \frac{\mathbb{E}\Big[X_A f(A)^2\Big]}{\mathbb{E}\Big[f(A)^2\Big]},$$

which is a minimum, as the second derivative $2\mathbb{E}\Big[f(A)^2\Big] \geq 0$ *almost surely. Plugging in the definition of* f *proves the claim.*