Monte Carlo Matrix Inversion

Complexity proof and implementation

Ahmed Touati

Reinforcement Learning class

Policy evaluation revisited

- Given a policy π, the value function V_π is the fixed point of Bellman equation: if d is the number of state, we have $V_\pi \in \mathbb{R}^d$

$$V_\pi = R_\pi + \gamma P_\pi V_\pi$$

  where
  - $R_\pi \in \mathbb{R}^d, R_\pi(s) = \mathbb{E}[R_t | S_t = s, A_{t:\infty} \sim \pi]$.
  - $P_\pi \in \mathbb{R}^{dxd}$ transition matrix: $(P_\pi)_{i,j} = \mathbb{P}[s_j | s_i]$
- The solution of Bellman equation is (complexity is d^3):

$$V_\pi^* = (I - \gamma P_\pi)^{-1} R_\pi$$

- Policy iteration: For each k:

$$V_\pi^{k+1} = R_\pi + \gamma P_\pi V_\pi^k$$

- By recursion

$$||V_\pi^* - V_\pi^k|| \leq \frac{1}{1-\gamma} ||V_\pi^* - V_\pi^0||$$

Policy evaluation revisited

- $$||V_\pi^* - V_\pi^k|| \leq \frac{1}{1-\gamma}||V_\pi^* - V_\pi^0||$$
- If ϵ is the amount of reduction desired i.e

  $$||V_\pi^* - V_\pi^k|| \leq \epsilon||V_\pi^* - V_\pi^0||$$

  then the number of multiplication required is:

  $$(1 + \frac{\log(\epsilon)}{\log(\gamma)})d^2$$
- It is better than exact method, but could we more improve the complexity?

Monte Carlo methods

Idea !!

A simple sum $\sum_k a_k$ could be interpreted as the expected value of random variable

$$\sum_k a_k = \sum_k \frac{a_k}{p_k} p_k = \mathbb{E}[Z]$$

where Z is random variable defined by $\mathbb{P}(Z = \frac{a_k}{p_k}) = p_k$ and $\{p_k\}$ a probability mass.

Neuman expansion of inverses

If $\rho(A) < 1$ than $(I - A)^{-1}$ exists and satisfies:
$(I - A)^{-1} = \lim_{N\to\infty} \sum_{n=0}^{N} A^n$

- $V = (I - \gamma P)R = R + \gamma PR + \gamma^2 P^2 R + ..$
- ith component:
  $V_i = R_i + \gamma \sum_{i_1} p_{ii_1} R_{i_1} + .. + \gamma^k \sum_{i_1 \ldots i_k} p_{ii_1} \ldots p_{i_{k-1} i_k} R_{i_k} + ..$

$$V_i = R_i + \sum_k \sum_{i_1 \ldots i_k} \gamma^k \prod_{j=1}^{k} p_{i_{j-1} i_j} R_{i_k}$$

Ulman and von-Neumann technique 1950

- Let's define a Markov chain with transition matrix $\tilde{P}$ and state set $\{1, 2, ..., d\}$.
- The chain starts in state i and is allowed o make k transitions.
- the chain's length k is a geometric distributed random variable with parameter p_{step} :

$$\mathbb{P}(k \text{ state transitions}) = p_{step}^k (1 - p_{step})$$

- Each trajectory starting in state i, $x_o = i \to x_1 = i_1 \to ... x_k = i_k$ corresponds to a unique term in the sum defining V_i
- For our case the RV Z (defined by $\mathbb{P}(Z = \frac{a_k}{p_k}) = p_k$) is:

$$\mathbb{P}(Z = \frac{\gamma^k \prod_{j=1}^k p_{i_{j-1} i_j} R_{i_k}}{p_{step}^k (1 - p_{step}) \prod_{j=1}^k \tilde{p}_{i_{j-1} i_j}}) = p_{step}^k (1 - p_{step}) \prod_{j=1}^k \tilde{p}_{i_{j-1} i_j}$$

Monte Carlo Method

- If we take $\tilde{P} = P$ and $p_{step} = \gamma$, we obtain:

$$\mathbb{P}(Z = \frac{R_{i_k}}{1-\gamma}) = \gamma^k (1-\gamma) \prod_{j=1}^{k} p_{i_{j-1} i_j}$$

- Our MC estimate is for a state i:

$$V_{MC}^n = R_i + \frac{1}{n} \sum_k^n Z_k$$

Complexity analysis

- work = number of multiplications required to achieve a given amount of reduction error ϵ
- ϵ is defined by:

$$|V^*(i) - V^n_{MC}| \leq \epsilon ||V^* - V^0_{MC}||$$

$$|V^*(i) - V^n_{MC}| \leq \epsilon ||(I - \gamma P)^{-1} R|| \leq \epsilon \frac{||R||}{1-\gamma}$$

- Z is bounded, $|Z| \leq \frac{||R||}{1-\gamma}$
- $\text{Var}(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 \leq (\frac{||R||}{1-\gamma})^2$
- If the distribution of Z is somewhat bell-shaped, the variance may be much less. Assume so that:

$$\text{Var}(Z) \leq \frac{1}{2}(\frac{||R||}{1-\gamma})^2$$

Central Limit Theorem

- As the expectation and the variance of Z are finite, we can apply the CLT theorem:

$$\frac{1}{\sqrt{\mathrm{Var}(\frac{\sum_k^n Z_k}{n})}}[\frac{\sum_k^n Z_k}{n} - \mathbb{E}[Z]] \Rightarrow N(0,1)$$

- $\mathrm{Var}(\frac{\sum_k^n Z_k}{n}) = \frac{\mathrm{Var}(Z)}{n}$
-

$$\mathbb{P}[|V^* - V^n_{MC}| \leq \epsilon \frac{||R||}{1-\gamma}] = \mathbb{P}[\frac{|\frac{\sum_k^n Z_k}{n} - \mathbb{E}[Z]|}{\sqrt{\mathrm{Var}(\frac{\sum_k^n Z_k}{n})}} \leq \epsilon \frac{\frac{||R||}{1-\gamma}}{\sqrt{\mathrm{Var}(\frac{\sum_k^n Z_k}{n})}}]$$

$$\leq \mathbb{P}[\frac{|\frac{\sum_k^n Z_k}{n} - \mathbb{E}[Z]|}{\sqrt{\mathrm{Var}(\frac{\sum_k^n Z_k}{n})}} \leq \frac{\epsilon\sqrt{n}}{\sqrt{2}}]$$

$$(=?) \rightarrow \int_{-\frac{\epsilon\sqrt{n}}{\sqrt{2}}}^{\frac{\epsilon\sqrt{n}}{\sqrt{2}}} \frac{1}{\sqrt{2\pi}} e^{\frac{-t^2}{2}}\,\mathrm{dt} = 1 - 2\int_{\frac{\epsilon\sqrt{n}}{\sqrt{2}}}^{\infty} \frac{1}{\sqrt{2\pi}} e^{\frac{t^2}{2}}\,\mathrm{dt}$$

- In order to obtain 95 % confidence level, $\frac{\epsilon\sqrt{n}}{\sqrt{2}}$ should be greater than 2. i.e $n \geq 1 + \frac{2}{\epsilon}$.
- Actually, in our derivation, n represents the number of Markov chain realizations (or trajectories). Each trajectory of length k requires k elementary operations. As the length is geometrically distributed with parameter γ, $\mathbb{E}[k] = \frac{1}{1-\gamma}$.
- At the end, the number of operations required to achieve ϵ error reduction is:

$$\text{work} = \frac{1}{1-\gamma}(1 + \frac{2}{\epsilon})$$

which proves the formula given in the article. ∎

Concentration inequalities

- The previous proof is based on the CLT which is an asymptotic result so the provided formula of the work is valid only when n is very large.
- We could give a better estimate of the work using the Hoeffding's inequality.

Hoeffding's inequality

Let $X_1, \ldots, X_n$ be independent random variables bounded by the intervals $[a_i, b_i] : a_i \leq X_i \leq b_i$. We define the empirical mean of these variables by $\overline{X} = \frac{1}{n}(X_1 + \cdots + X_n)$. Then,

$$\mathbb{P}\left(|\overline{X} - \mathbb{E}\left[\overline{X}\right]| \geq t\right) \leq 2\exp\left(-\frac{2n^2t^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right)$$

and if $b_i - a_i \leq C$, then:

$$\mathbb{P}\left(|\overline{X} - \mathbb{E}\left[\overline{X}\right]| \geq t\right) \leq 2\exp\left(-\frac{2nt^2}{C^2}\right)$$

Better estimate of the work

- Let's apply Hoeffding's inequality: In our case $C = \frac{||R||}{1-\gamma}$.

$$\mathbb{P}[|V^* - V_{MC}^n| \geq \epsilon \frac{||R||}{1-\gamma}] \leq 2 \exp\left(-\frac{2n(\epsilon\frac{||R||}{1-\gamma})^2}{C^2}\right)$$
$$= 2 \exp\left(-2n\epsilon^2\right)$$

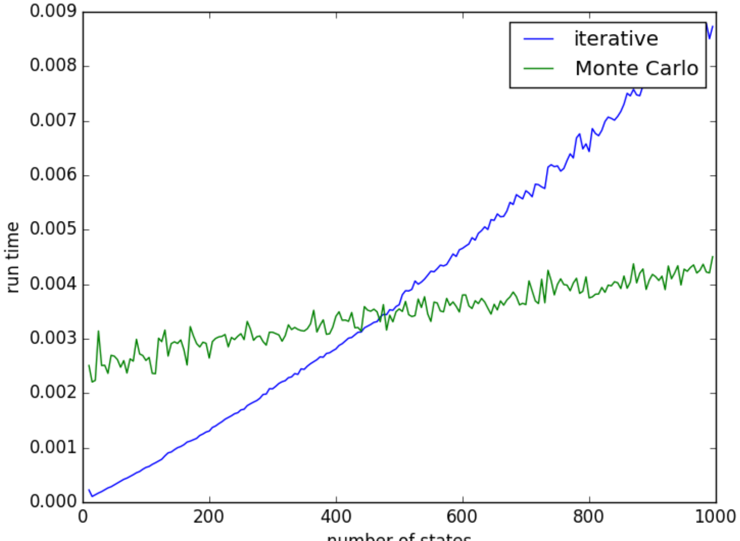
- In order to obtain 95 % confidence level, $n \geq 1 - \log(\frac{1-0.95}{2})\frac{1}{\epsilon^2}$
- Finally:

$$work = \frac{1}{1-\gamma}(1 + \frac{3.68}{\epsilon^2})$$

■

Let's check experimentally our beautiful formulas

$$\text{work}_{\text{Monte-Carlo}} = \frac{1}{1-\gamma}\left(1 + \frac{3.68}{\epsilon^2}\right) \longleftrightarrow \text{work}_{\text{iterative}} = \left(1 + \frac{\log(\epsilon)}{\log(\gamma)}\right) d^2$$



Questions?