Multi-Agent Systems

Homework Assignment 4

MSc AI, VU

E.J. Pauwels

Version: November 26, 2021 — Deadline: Friday, December 3, 2021 (23h59)

4 Monte Carlo simulation

4.1 MC sampling

Recall that Monte Carlo sampling allows us to estimate the expectation of a random function by sampling from the corresponding probability distribution. More precisely, if $f(x)$ is a 1-dim (continuous) probability density, and $X \sim f$ is a stochastic variable distributed according to this density f, then the expected value of some function φ can be estimated using Monte Carlo sampling by:

$$E_f(\varphi(X)) \equiv \int \varphi(x)f(x)\,dx \approx \frac{1}{n}\sum_{i=1}^{n}\varphi(X_i) \qquad \text{for sample of independent } X_1, X_2, \ldots, X_n \sim f.$$

1. Assume that $X \sim N(0,1)$ is standard normal. Estimate the mean value $E(\cos^2(X))$. Quantify the uncertainty on your result.

2. Suppose you're designing a deep neural network that needs to maximize some score function S. The actual design of the network depends on some hyperparameter A. Training the networks is computationally very demanding and time consuming, and as a consequence you have only been able to perform ten experiments to date. Based on these ten data points you observe a slight positive correlation of 0.3 between the value of the hyperparameter A and the score S. If this result is genuine, it suggest to increase A in the next experiment in order to improve the score. But if the correlation is not significant, increasing A could lead you astray. How would you use MC to decide whether the correlation is significant?
Hint: Compute the empirical p-value of the observed result, under the assumption of independence.

4.2 Importance Sampling

Importance sampling extends the basic MC approach to cases where it is difficult to sample from f but (relatively) easy to sample from a (somewhat) similar distribution g. More precisely:

$$
\begin{aligned}
E_f(\varphi(X)) &= \int \varphi(x) f(x)\, dx \\
&= \int \varphi(x) \frac{f(x)}{g(x)}\, g(x)\, dx \equiv E_g\left[\varphi(X)\frac{f(X)}{g(X)}\right] \\
&\approx \frac{1}{n}\sum_{i=1}^{n} \varphi(X_i)\frac{f(X_i)}{g(X_i)} \qquad \text{for sample of independent } X_1, \ldots, X_n \sim g.
\end{aligned}
$$

1. Let $X \sim N(0,1)$ be a standard normal stochastic variable. Use importance sampling to estimate $E(X^2)$ by sampling from a uniform distribution $q \sim U(-5,5)$ on the interval $[-5,5]$. What value do you expect (based on your knowledge of the normal distribution)? How accurate is your estimate based on importance sampling?

2. Suppose some random process produces output ($-1 \le X \le 1$) that is distributed according to the following continuous density:

$$f(x) = \frac{1+\cos(\pi x)}{2} \qquad (\text{for } -1 \le x \le 1).$$

Again we are interested in estimation $E(X^2)$. However, as this is not a standard distribution it makes sense to use importance sampling to estimate this value. Quantify the uncertainty on your result.

4.3 Kullback-Leibler divergence

The Kullback-Leibler (KL) divergence quantifies the similarity (or dissimilarity) of two probability densities. More specifically, given two continuous (1-dim) probability densities f, g, the KL-divergence is defined as:

$$KL(f||g) = \int_{-\infty}^{\infty} f(x) \log\left(\frac{f(x)}{g(x)}\right) dx \quad \equiv \quad E_f\left[\log\left(\frac{f(X)}{g(X)}\right)\right] \tag{1}$$

1. Let $f \sim N(\mu, \sigma^2)$ and $g \sim N(\nu, \tau^2)$ both be normal distributions. Express $KL(f||g)$ as a function of the means and variances of f and g. We mention in passing that the KL expression in eq.1 is called a **divergence** rather than a **distance** because it's not symmetric. Use the expression obtained above to convince yourself of this fact.

2. Check your theoretical result in (1) by computing a sample-based estimate of the KL-divergence (Monte Carlo simulation). Pick an appropriate sample size. Compare the MC estimate to the theoretical result.

5 Exploitation versus Exploration

5.1 UCB versus ϵ-greedy for k-bandit problem

Write a programme to experiment with the exploration/exploitation for the k-bandit problem (pick some value $5 \le k \le 20$). Assume that the arms (a) generate normally distributed rewards with unit standard deviation, but different means $q(a)$ (e.g. randomly generated). Assume that in every single experiment the agent can take a total of $T = 1000$ actions (i.e. arm pulls). Let $L(t)$ be the expected total regret at time t, defined as:

$$\ell(t) = E\left(\sum_{i=1}^{t}(q^* - q(a_i))\right)$$

- Compute the experimental $L(t)$ curves for different strategies (ϵ-greedy for different values of ϵ, UCB). Compare to the theoretical lower bound found by Lai-Robbins:

$$\ell(t) \geq A \log(t) \qquad \text{where} \qquad A = \sum_{a:\Delta_a \neq 0} \frac{\Delta_a}{KL(f_a || f_a^*)} \quad \text{and} \quad \Delta_a = q^* - q(a).$$

the arm with the highest return?

- Compute and compare the percentage correct decisions (selection of correct arm) under the different strategies (i.e. ϵ-greedy for different values of ϵ, UCB). What is the influence of the UCB hyper-parameter c?

PS: *No need to submit code, only the results.*