Solutions for Quiz I Practice Problems

26/01/2023

Lecturer: Dr. S.S. Peruru *Scribe: Swastik Sharma (21104286) & Himanshu Sood (190381)*

Exercise 4

In the class, we have looked at Thompson sampling for the Gaussian case, where the reward distributions are $\{\mathcal{N}(\mu(a), 1)\}_{a\in\mathcal{A}}$, and the prior distributions for all the arms is $\mathcal{N}(0, 1)$. Now, assume that instead of the standard Gaussian $\mathcal{N}(0, 1)$ as a prior, we have the following Gaussian distributions as priors $\{\mathcal{N}\left(v_0(a), \sigma_0(a)\right)\}_{a\in\mathcal{A}}$

1. Prove that the posterior of an arm a after observing 1 reward sample r from reward distribution $\mathcal{N}(\mu(a), 1)$ is given by

$$\mathcal{N}\left(\frac{v_0(a) + (\sigma_0(a))^2\, r}{1 + (\sigma_0(a))^2}, \frac{(\sigma_0(a))^2}{1 + (\sigma_0(a))^2}\right)$$

2. Using the above result, write the Thompson sampling algorithm for this case.

Solution:

Given,

$$\text{Prior} : \{\mathcal{N}\left(v_0(a), \sigma_0(a)\right)\}_{a\in\mathcal{A}}$$

$$\text{Rewards} : \{\mathcal{N}(\mu(a), 1)\}_{a\in\mathcal{A}}$$

We know from Baye's Theorem that:

$$\mathbf{Pr}(\theta|r) \propto \mathbf{Pr}(r|\theta) + \mathbf{Pr}(\theta) \tag{1}$$

Where r is one sample from the reward distribution and θ is a random variable sampled from the prior distribution, over which we are updating our belief.

Now, we know for Gaussian Distribution R.H.S. of (1) can be written as,

$$
\begin{aligned}
\mathbf{Pr}(\theta|r) &\propto \mathbf{Pr}(r|\theta) + \mathbf{Pr}(\theta) \\
&\propto \exp\left\{-\frac{1}{2}((r-\theta))^2\right\} \cdot \exp\left\{-\frac{1}{2(\sigma_0(a))^2}\left(\theta - v_0(a)\right)^2\right\} \\
&\propto \exp - \left\{\frac{((r-\theta))^2}{2} + \frac{\left(\theta - v_0(a)\right)^2}{2(\sigma_0(a))^2}\right\} \\
&\propto \exp - \left\{\frac{\theta^2}{2} - r\theta + \frac{\left(\theta\right)^2}{2(\sigma_0(a))^2} - \frac{\theta v_0(a)}{2(\sigma_0(a))^2}\right\} \\
&\propto \exp - \left\{\frac{\theta^2}{2}\frac{\left(\sigma_0(a)\right)^2 + 1}{(\sigma_0(a))^2} - \theta(r + \frac{v_0(a)}{(\sigma_0(a))^2}\right\} \\
&\propto \exp - \frac{1}{2\left(\sigma_0(a)\right)^2}\left\{\theta^2((\sigma_0(a))^2 + 1) - 2\theta(r(\sigma_0(a))^2 + v_0(a))\right\} \\
&\propto \exp - \frac{((\sigma_0(a))^2 + 1)}{2\left(\sigma_0(a)\right)^2}\left\{\theta^2 - 2\theta\frac{(r(\sigma_0(a))^2 + v_0(a))}{((\sigma_0(a))^2 + 1)}\right\} \\
&\propto \exp - \frac{((\sigma_0(a))^2 + 1)}{2\left(\sigma_0(a)\right)^2}\left\{\left(\theta - \frac{(r(\sigma_0(a))^2 + v_0(a))}{((\sigma_0(a))^2 + 1)}\right)^2\right\}
\end{aligned}
$$

The last expression can be expressed as:

$$
\mathbf{Pr}(\theta|r) = \mathcal{N}\left(\frac{(r(\sigma_0(a))^2 + v_0(a))}{((\sigma_0(a))^2 + 1)}, \frac{(\sigma_0(a))^2}{((\sigma_0(a))^2 + 1)}\right) \tag{2}
$$

Hence Proved.

The Thompson Sampling algorithm for this case is represented in Algorithm 1.

Algorithm 1 Thompson Sampling for Gaussian Priors

Set $\mu_0(a) = 0 \; \forall \, a \in A$
for $t \geq 1$ **do**
  for `each arm a` **do**
    Sample $\widetilde{\theta}_t(a)$ from $\mathcal{N}\left(\frac{(r(\sigma_0(a))^2 + v_0(a))}{((\sigma_0(a))^2+1)}, \frac{(\sigma_0(a))^2}{((\sigma_0(a))^2+1)}\right)$
  end for
  Play $a(t)$ = $\underset{a}{\operatorname{argmax}}\, \widetilde{\theta}_t(a)$
  Observe reward r_t
  if $a(t) = a$ **then**
    Update $\bar{\mu}_t(a)$ based on observed reward r_t
  end if
end for

Exercise 5

Consider two Gaussian distributions $\mathcal{N}(\mu_a, \sigma)\,\mathcal{N}(\mu_b, \sigma)$. Prove that the KL divergence between these two distributions is $\frac{1}{2\sigma^2}\left(\mu_a - \mu_b\right)^2$.

Solution:

Given two distributions, let them be named:

$$P = \mathcal{N}(\mu_a, \sigma)$$

$$Q = \mathcal{N}(\mu_b, \sigma)$$

As they are normal distributions they can be defined as:

$$P(x) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp - \left\{ \frac{(x-\mu_a)^2}{2\sigma^2} \right\}$$

$$Q(x) \triangleq \frac{1}{\sqrt{2\pi\sigma^2}} \cdot \exp - \left\{ \frac{(x-\mu_b)^2}{2\sigma^2} \right\}$$

From the KL-Divergence of two distributions, we know that:

$$\begin{aligned}
KL(P||Q) &= \int_x P(x) \cdot \log\left(\frac{P(x)}{Q(x)}\right) dx \\
&= \int_x P(x) \cdot \log\left(\exp - \left\{\frac{(x-\mu_a)^2 - (x-\mu_b)^2}{2\sigma^2}\right\}\right) dx \\
&= \int_x P(x) \cdot \left(\frac{(x-\mu_b)^2 - (x-\mu_a)^2}{2\sigma^2}\right) dx \\
&= \frac{1}{2\sigma^2}\left\{-\int_x P(x)(x-\mu_a)^2 dx + \int_x P(x)(x-\mu_b)^2 dx\right\} \\
&= \frac{1}{2\sigma^2}\left\{-\sigma^2 + \int_x P(x)(x-\mu_b+\mu_a-\mu_a)^2 dx\right\} \\
&\qquad\qquad\qquad \text{(Beacuse } \textstyle\int_x P(x)(x-\mu_a)^2 dx = \sigma^2 \text{ and adding and subtracting } \mu_a\text{)} \\
&= \frac{1}{2\sigma^2}\left\{-\sigma^2 + \int_x P(x)(x-\mu_a)^2 dx + \int_x P(x)(\mu_a-\mu_b)^2 dx + 2(\mu_a-\mu_b)\int_x P(x)(x-\mu_a) dx\right\} \\
&= \frac{1}{2\sigma^2}\left\{-\sigma^2 + \sigma^2 + (\mu_a-\mu_b)^2\right\} \\
&\quad \text{(Beacuse } \textstyle\int_x P(x)(x-\mu_a)^2 dx = \sigma^2, \int_x P(x)(x-\mu_a) dx = 0 \text{ and } \int_x P(x) dx = 1\text{)} \\
&= \frac{(\mu_a-\mu_b)^2}{2\sigma^2}
\end{aligned}$$

Hence proved that KL divergence between $\mathcal{N}(\mu_a, \sigma)\,\mathcal{N}(\mu_b, \sigma)$ is $\frac{1}{2\sigma^2}\left(\mu_a - \mu_b\right)^2$.

Exercise 6

Recall the hypothesis testing problem with two distributions $\mathcal{N}(0,1), \mathcal{N}(\Delta,1)$. We are given T samples from one of these two distributions, and we have to predict from which of these two distributions the samples were drawn. Assume $\Delta = \frac{1}{\sqrt{T}}$. Use the above theorem to show that the prediction of hypothesis testing can go wrong with a constant probability (i.e., the constant does not depend on T).

Solution:

Given that there are two distributions $\mathcal{N}(0,1), \mathcal{N}(\Delta,1)$ and that $\Delta = \frac{1}{\sqrt{T}}$ where T is the number of samples.
We know that after taking N samples, the distributions are given by, let these be named P and Q respectively:

$$\mathcal{N}\left(0, \frac{1}{N}\right) = P$$

$$\mathcal{N}\left(\Delta, \frac{1}{N}\right) = Q$$

From the KL-Divergence of two distributions, we know that;

$$\begin{aligned}
KL(P||Q) &= \int_x P(x) \cdot \log\left(\frac{P(x)}{Q(x)}\right) dx \\
&= \int_x P(x) \cdot \log\left(\frac{\exp\left\{-\frac{N}{2} \cdot (x^2)\right\}}{\exp\left\{-\frac{N}{2} \cdot (x-\Delta)^2\right\}}\right) dx \\
&= \int_x P(x) \cdot \log\left(\exp\left\{-\frac{N}{2}\left[x^2 - (x-\Delta)^2\right]\right\}\right) dx \\
&= -\frac{N}{2} \int_x P(x) \cdot \left[x^2 - (x-\Delta)^2\right] dx \\
&= -\frac{N}{2} + \frac{N}{2} \int_x P(x) \cdot x^2 + \frac{N}{2} \int_x P(x) \cdot \Delta^2 - \frac{N}{2} \int_x P(x) \cdot 2x\Delta dx \\
&= \frac{N}{2} \int_x P(x) \cdot \Delta^2 dx
\end{aligned}$$

Therefore, as $\int_x P(x)dx = 1$, the KL divergence of the distributions P and Q can be written as-

$$KL(P||Q) = \frac{\Delta^2 \cdot N}{2} \tag{3}$$

We know from the Bretagnolle - Huber Inequality, for 2 distributions P & Q on the same sample space, there for any event A, we have:

$$P(A^c) + Q(A) \geq \frac{1}{2} \exp \{-KL(P, Q)\} \, \forall a \in A \tag{4}$$

Here A is the event in which sample mean $\leq a > \frac{\Delta}{2}$, and
A^c is the complement of that event with sample mean $\geq a > \frac{\Delta}{2}$

Here, $P(A^c)$ signifies the probability of the wrong predictions from samples from distribution P and $Q(A)$ signifies the probability of the wrong predictions from samples from distribution Q

From (3) and (4) we get:

$$P(A^c) + Q(A) \geq \frac{1}{2} \exp \left\{-\frac{\Delta^2 \cdot N}{2}\right\} \forall a \in A \tag{5}$$

As $\Delta = \frac{1}{\sqrt{T}}$ or $\frac{1}{\sqrt{N}}$, equating in (5) yields:

$$P(A^c) + Q(A) \geq \frac{1}{2} \exp \left\{-\frac{1}{2}\right\}$$

$$P(A^c) + Q(A) \geq 0.3033 \tag{6}$$

Therefore as can be seen from the analysis that the prediction of hypothesis testing can go wrong with a constant probability (≥ 0.3033)