02477 Bayesian Machine Learning 2024: Assignment 1

This is the first assignment out of three in the Bayesian machine learning course 2024. The assignment is a group work of 3-5 students (make your own groups) and hand in via DTU Learn). The assignment is **mandatory**. The deadline is **25th of February 23:59**.

Being able to manipulate probability distributions is crucial for probabilistic machine learning. The purpose of this exercise is to become more familiar with Bayesian computation and recap central concepts of probability theory, e.g.

- simple moment calculations
- the sum rule for marginalization and the product rule constructing joint distributions
- computing the analytical expression for conditional distributions of simple Bayesian models
- manipulating Gaussian distributions

You will also practice manipulation and updating knowledge based on data using Bayesian inference. The assignment contains 4 parts and a total of 24 questions.

Part 1: The beta-binomial model

Your friend has set up a website for her new business. So far $N = 17$ potential customers has visited her site, but only $y = 1$ has completed a purchase. To plan her future investments, she asks you for help to compute the probability that at least one of the next 20 customers will make a purchase. You decide to model the problem using the beta-binomial model with a uniform prior distribution on the probability of making a purchase $\mu \in [0, 1]$:

$$\mu \sim \text{Beta}(a_0, b_0), \tag{1}$$

$$y|\mu \sim \text{Binomial}(N, \mu) \tag{2}$$

where $a_0 = b_0 = 1$.

Task 1.1: Argue that the posterior distribution of $\boldsymbol{\mu}$ given $\boldsymbol{y = 1}$ is a $\boldsymbol{p(\mu|y = 1) = \textbf{Beta}(2, 17)}$.

Task 1.2: Compute the posterior mean and 95% posterior interval for $\boldsymbol{\mu}$

Hint: See Section 4.6.6 in Murphy1 for details on posterior/credibility intervals

Your friend is not really happy with the relatively large uncertainty for your reported estimate, so she decides to collect more data and it turns out that $y_2 = 4$ of the next $N_2 = 20$ customers decides to make a purchase.

Task 1.3: Update your knowledge of the purchase probability μ with the new data using the posterior distribution from the previous task as a prior. Recompute the posterior mean and 95% interval.

Task 1.4: Consider a random variable μ, where the outcome space is the unit interval and the log probability density function is given by $\ln p(\mu) = 95 \ln \mu + 10 \ln(1 - \mu) + c$, where $c \in \mathbb{R}$ is a constant independent of μ. Identify the type of distribution (e.g. Gaussian, normal, binomial, Bernoulli etc.) and compute the mean of μ.

Hint: Can you re-write the log density into a known functional form? After you identified the distribution, computing the mean should be easy.

Part 2: The sum and product rule for a simple toy model

Suppose you have a strange machine with a button and a display in your lab. The machine has a single button, which can be either turned on ($x = 1$) or turned off ($x = 0$). The machine outputs numbers of the form

$$y = \begin{cases} 2 + \epsilon & \text{when } x = 1, \\ 0 + \epsilon & \text{when } x = 0 \end{cases} \tag{3}$$

where $\epsilon \sim \mathcal{N}(0, \sigma_y)^2$ is *independent* Gaussian noise. Assuming $x \in \{0, 1\}$, we can write this more compactly:

$$y = 2x + \epsilon, \tag{4}$$

or equivalently,

$$p(y|x) = \mathcal{N}(y|2x, \sigma_y^2). \tag{5}$$

This is the *likelihood* of the model. Assume the probability of finding the machine in a state, where the button is turned on is $p(x = 1) = \alpha$, where $\alpha \in [0, 1]$. That is,

$$x = \begin{cases} 0 & \text{with probability } 1 - \alpha \\ 1 & \text{with probability } \alpha \end{cases}, \tag{6}$$

or equivalently,

$$p(x) = \text{Bernoulli}(x|\alpha). \tag{7}$$

This is the *prior distribution* of the model. We typically refer to α and σ_y^2 as *hyperparameters* of the model.

Task 2.1: Show that the prior mean of x is $\mathbb{E}[x] = \alpha$

Hint: Use definition of expectation for discrete random variables

Task 2.2: Use the product rule to write up the joint distribution for $p(y, x)$

Hint: This is meant to be an easy question: you don't need to calculate anything.

Task 2.3: Use the sum rule to compute the marginal distribution of $p(y)$

Hint: See eq. (1.10) in Bishop

Task 2.4: Compute the mean of the distribution of $p(y)$

Hints: this exercise can be solved in several ways, e.g. either directly using eq. (1.34) in Bishop or using linearity of the expectation operator.

Task 2.5: Compute the second moment the distribution of $p(y)$

Hints:

- *This can also be solved in several ways, e.g. either using eq. (1.34) in Bishop again using $f(y) = y^2$ or using linearity of the expectation operator.*
- *Recall the second moment of Gaussian random variable $y \sim \mathcal{N}(m, v)$ is $\mathbb{E}\left[y^2\right] = m^2 + v$.*

Task 2.6: Compute the variance of $p(y)$

Hints: The variance is easily computed using the first and second moment: $\mathbb{V}[y] = \mathbb{E}\left[y^2\right] - \mathbb{E}[y]^2$.

Task 2.7: Use Bayes rule to show that the posterior distribution $p(x = 1|y)$ is given by

$$p(x=1|y) = \frac{1}{1 + \frac{(1-\alpha)\mathcal{N}(y|0,\sigma_y^2)}{\alpha\mathcal{N}(y|2,\sigma_y^2)}}. \tag{8}$$

Hints: Write up Bayes rule and insert the expressions for the prior, likelihood and marginal likelihood and divide all terms by the numerator.

Task 2.8: Assume the machine outputs the value $y = 1.5$. What is the posterior probability that $x = 1$ assuming $\sigma_y^2 = \frac{1}{2}$ and $\alpha = \frac{1}{2}$? What if the noise variance is $\sigma_y^2 = 5$ instead?

Part 3: A simple Linear Gaussian system

We will now assume the button on the machine has been replaced with a dial that represents a real number, i.e. $x \in \mathbb{R}$, instead of a binary variable. We assume the same likelihood

$$y = 2x + \epsilon, \tag{9}$$

but now $x \in \mathbb{R}$ is now a real number and $\epsilon \sim \mathcal{N}(0, \sigma_y^2)$ is independent Gaussian noise. Assuming a Gaussian prior for x, we can write the complete probabilistic model as follows

$$p(x) = \mathcal{N}(x|m_x, \sigma_x^2) \tag{10}$$
$$p(y|x) = \mathcal{N}(y|2x, \sigma_y^2), \tag{11}$$

where $m_x \in \mathbb{R}$ and $\sigma_x^2 > 0$ are the *prior mean and variance*, respectively, of x.

Task 3.1: Use Bayes rule to show that

$$\log p(x|y) = \log \mathcal{N}(y|2x, \sigma_y^2) + \log \mathcal{N}(x|m_x, \sigma_x^2) + K, \tag{12}$$

where K is a constant independent of x.

Hints: Write up Bayes rule and take the logarithm on both sides

Task 3.2: Show that

$$\log p(x|y) = -\frac{(y-2x)^2}{2\sigma_y^2} - \frac{(x-m_x)^2}{2\sigma_x^2} + K_1, \tag{13}$$

where K_1 is a constant independent of x.

Task 3.3: Show that eq. (13) can be expressed as

Hints: Start by expanding the parentheses in eq. (13) and then factor out $-\frac{1}{2}x^2$ and x

$$\log p(x|y) = -\frac{1}{2}x^2\left(\frac{2^2}{\sigma_y^2} + \frac{1}{\sigma_x^2}\right) + x\left(\frac{2y}{\sigma_y^2} + \frac{m_x}{\sigma_x^2}\right) + K_2, \tag{14}$$

where K_2 is a constant independent of x.

Task 3.4: Argue that the distribution $p(x|y)$ must be Gaussian density, i.e. $p(x|y) = \mathcal{N}(x|m, v)$ for some m and v.

Task 3.5: Show that the variance is given by

$$v^{-1} = \frac{2^2}{\sigma_y^2} + \frac{1}{\sigma_x^2} \tag{15}$$

Task 3.6: Show that the mean is given by

$$m = \frac{2}{2^2 + \frac{\sigma_y^2}{\sigma_x^2}} y + \frac{1}{\frac{2^2\sigma_x^2}{\sigma_y^2} + 1} m_x \tag{16}$$

Task 3.7: What happens to the posterior mean when σ_x^2 is fixed and $\sigma_y^2 \to \infty$?

Task 3.8: What happens to the posterior mean when σ_y^2 is fixed and $\sigma_x^2 \to \infty$?

Part 4: Bayesian inference for two-parameter model

Consider a toy dataset with 4 observations $\boldsymbol{y} = \{1, 2, 3, 4\}$, where $y_i \in \mathbb{R}$ denotes the i'th observation.

Consider the following model for the data

$$\begin{aligned} y_i &\sim \mathcal{N}(\mu, \sigma^2) && \text{(likelihood)} \\ \mu &\sim \mathcal{N}(0, 10) && \text{(prior for } \mu\text{)} \\ \sigma^2 &\sim \text{Inv-Gamma}(1, 1) && \text{(prior for } \sigma^2\text{)} \end{aligned}$$

The product rule yields the joint distribution

$$p(\mathbf{y}, \mu, \sigma^2) = p(\boldsymbol{y}|\mu, \sigma^2)\mathcal{N}(\mu|0, 10)\text{Inv-Gamma}(\sigma^2|1, 1) \tag{17}$$

$$= \prod_{i=1}^{4} p(y_i|\mu, \sigma^2)\mathcal{N}(\mu|0, 10)\text{Inv-Gamma}(\sigma^2|1, 1) \tag{18}$$

Task 4.1: Implement a function for evaluating the log prior, the likelihood and the joint distribution and make a contour plots of the prior, the likelihood and the posterior distribution for $\mu \in [-5, 5]$ and $\sigma^2 \in \left[10^{-6}, 5\right]$.

Hints: Implementing the prior, likelihood and joint should be the most demanding component of this part of the exercise. You should be able to re-use the code from exercise for plotting with minor modifications.

Task 4.2: Use a grid approximation to compute and plot the approximate marginal posterior distributions $p(\mu|\boldsymbol{y})$ and $p(\sigma^2|\boldsymbol{y})$

Hints: You can re-use the code from exercise 2, specifically the GridApproximation2D class.

Task 4.3: Compute and report the posterior mean and an approximate 95% credibility interval for μ and σ^2

Task 4.4: Compute and plot the posterior predictive density for a new observation y^* using: 1) the plugin approximation with MLE, 2) the plugin approximation using MAP, and 3) the grid approximation

Hints: You can use the Grid2D class for identifying the MAP and MLE. For computing the full posterior predictive density, you need to compute the predictive likelihood $p(y^*|\mu, \sigma^2)$ for all combinations in the grid and the compute a weighted sum wrt. the posterior distribution.