Machine Learning (SS 23)

Assignment 3: Bayes (Solution)

Mojtaba Nayyeri
Mojtaba.Nayyeri@ipvs.uni-stuttgart.de

Akram Sadat Hosseini
Akram.Hosseini@ipvs.uni-stuttgart.de

Nadeen Fathallah
Nadeen.Fathallah@ipvs.uni-stuttgart.de

Rodrigo Lopez Portillo Alcocer
rodrigo.lopez-portillo-alcocer@ipvs.uni-stuttgart.de

Tim Schneider
timphillip.schneider@ipvs.uni-stuttgart.de

Osama Mohammed
osama.mohammed@ipvs.uni-stuttgart.de

Daniel Frank
daniel.frank@ipvs.uni-stuttgart.de

This assignment sheet consists of 4 pages, plus the 2 tasks contained in the .ipynb notebook.

Submit your solution in ILIAS as a single PDF file.[1] Make sure to list full names of all participants, matriculation number, study program and B.Sc. or M.Sc on the first page. Optionally, you can *additionally* upload source files (e.g. PPTX files). If you have any questions, feel free to ask them in the exercise forum in ILIAS.

Submission is open until Monday, 15th May 2023, 12:00 noon.

[1] Your drawing software probably allows to export as PDF. An alternative option is to use a PDF printer. If you create multiple PDF files, use a merging tool (like pdfarranger) to combine the PDFs into a single file.

Maximum Likelihood vs. Bayes

An unfair coin is tossed seven times and the event (head or tail) is recorded at each iteration. The observed sequence of events is $\mathcal{D} = (x_1, x_2, \ldots, x_7) = (\text{ head, head, tail, tail, head, head, head })$. We assume that all tosses $x_1, x_2, \ldots$ have been generated independently following the Bernoulli probability distribution

$$P(x \mid \theta) = \begin{cases} \theta & \text{if} \quad x = \text{ head} \\ 1-\theta & \text{if} \quad x = \text{ tail} \end{cases}$$

where $\theta \in [0, 1]$ is an unknown parameter.

1. State the likelihood function $P(\mathcal{D} \mid \theta)$, that depends on the parameter θ.

   Solution Since the coin tosses are independent events, we can express the likelihood function as the product of the probabilities for each individual toss. In this case, the likelihood function $P(\mathcal{D} \mid \theta)$ can be written as:

$$\begin{aligned} P(\mathcal{D} \mid \theta) &= P(x_1, x_2, \ldots, x_7 \mid \theta) \\ &= P(x_1 \mid \theta) \cdot P(x_2 \mid \theta) \cdots P(x_7 \mid \theta) \\ &= \theta^{\text{number of heads}} \cdot (1-\theta)^{\text{number of tails}} \end{aligned}$$

   Given the observed sequence of events $\mathcal{D} =$ head, head, tail, tail, head, head, head, we have 5 heads and 2 tails. Therefore, the likelihood function is:

$$P(\mathcal{D} \mid \theta) = \theta^5 \cdot (1-\theta)^2$$

2. Compute the maximum likelihood solution $\hat{\theta}$, and evaluate for this parameter the probability that the next two tosses are "head", that is, evaluate $P\left(x_8 = \text{ head, } x_9 = \text{ head } \mid \hat{\theta}\right)$.

   Solution To compute the maximum likelihood solution $\hat{\theta}$, we need to maximize the likelihood function $P(\mathcal{D} \mid \theta)$. Taking the logarithm of the likelihood function and differentiating with respect to θ, we get:

$$\begin{aligned} \frac{d}{d\theta} \log P(\mathcal{D} \mid \theta) &= \frac{d}{d\theta} [5 \log \theta + 2 \log(1-\theta)] \\ &= \frac{5}{\theta} - \frac{2}{1-\theta} \end{aligned}$$

   Setting the derivative to zero and solving for θ, we find the maximum likelihood estimate $\hat{\theta}$:

$$\begin{aligned} \frac{5}{\hat{\theta}} - \frac{2}{1-\hat{\theta}} &= 0 \\ 5(1-\hat{\theta}) &= 2\hat{\theta} \\ \hat{\theta} &= \frac{5}{7} \end{aligned}$$

   Now, we want to evaluate the probability that the next two tosses are heads, given the maximum likelihood estimate $\hat{\theta}$. Using the Bernoulli probability distribution, we can compute this probability:

$$\begin{aligned} P(x_8 = \text{head}, x_9 = \text{head} \mid \hat{\theta}) &= P(x_8 = \text{head} \mid \hat{\theta}) \cdot P(x_9 = \text{head} \mid \hat{\theta}) \\ &= \hat{\theta}^2 \\ &= \left(\frac{5}{7}\right)^2 \\ &= \frac{25}{49} \end{aligned}$$

Therefore, the probability that the next two tosses are heads given the maximum likelihood estimate $\hat{\theta}$ is $\frac{25}{49}$.

3. We now adopt a Bayesian view on this problem, where we assume a prior distribution for the parameter θ defined as:

$$p(\theta) = \begin{cases} 1 & \text{if } 0 \leq \theta \leq 1 \\ 0 & \text{else} \end{cases}$$

Compute the posterior distribution $p(\theta \mid \mathcal{D})$, and evaluate the probability that the next two tosses are head, that is,

$$\int P\left(x_8 = \text{ head }, x_9 = \text{ head } \mid \theta\right) p(\theta \mid \mathcal{D}) d\theta$$

Solution To compute the posterior distribution $p(\theta \mid \mathcal{D})$, we will use Bayes' theorem:

$$p(\theta \mid \mathcal{D}) = \frac{P(\mathcal{D} \mid \theta) p(\theta)}{P(\mathcal{D})}$$

Since our prior $p(\theta)$ is a uniform distribution over $[0, 1]$, we have:

$$p(\theta) = \begin{cases} 1 & \text{if } 0 \leq \theta \leq 1 \\ 0 & \text{else} \end{cases}$$

Therefore, we can write the unnormalized posterior distribution as:

$$\begin{aligned} p(\theta \mid \mathcal{D}) &\propto P(\mathcal{D} \mid \theta) p(\theta) \\ &= \theta^5 (1 - \theta)^2 \end{aligned}$$

To normalize the posterior distribution, we need to find the normalization constant Z, which can be computed by integrating the unnormalized posterior distribution over the parameter space:

$$\begin{aligned} Z &= \int_0^1 \theta^5 (1 - \theta)^2 d\theta \\ &= \int_0^1 \theta^2 - 2\theta^6 + \theta^5 d\theta \\ &= \int_0^1 \theta^7 d\theta - 2 \int_0^1 \theta^6 d\theta + \int_0^1 \theta^5 d\theta \\ &= \frac{1}{8} - 2 \cdot \frac{1}{7} + \frac{1}{6} \end{aligned}$$

$$= \frac{1}{168}$$

Thus, the normalized posterior distribution is:

$$p(\theta \mid \mathcal{D}) = \frac{1}{168}\theta^5(1-\theta)^2$$

Now, we want to evaluate the probability that the next two tosses are heads. Using the expression for $P(x_8 = \text{head}, x_9 = \text{head} \mid \theta)$ and the posterior distribution $p(\theta \mid \mathcal{D})$, we can compute the probability:

$$\begin{aligned}
&\int P(x_8 = \text{head}, x_9 = \text{head} \mid \theta) p(\theta \mid \mathcal{D}) d\theta \\
&= \int \theta^2 \cdot \frac{1}{168}\theta^5(1-\theta)^2 d\theta \\
&= \frac{1}{168}\int_0^1 \theta^7(1-\theta)^2 d\theta \\
&= \frac{7}{15}
\end{aligned}$$

Therefore, the probability that the next two tosses are heads given the Bayesian posterior distribution is $\frac{7}{17}$.