Machine Learning I at TU Berlin

Assignment 2 - Group PTHGL

October 28, 2018

Exercise 1

(a)

The variables x and y are statistically independent if the following condition is fulfilled

$$p(x,y) = p(x) \cdot p(y) \qquad , \forall (x,y) \in \mathbb{R}_+^2 \quad .$$

To calculate $p(x)$ we have to integrate the given probability distribution $p(x,y) = \lambda\eta e^{-\lambda x - \eta y}$ over the y and vice versa.

$$p(x) = \int_0^\infty \lambda\eta e^{-\lambda x - \eta y} \mathrm{d}y = \left[-\lambda e^{-\lambda x - \lambda y}\right]_0^\infty = \lambda e^{-\lambda x}$$
$$p(y) = \int_0^\infty \lambda\eta e^{-\lambda x - \eta y} \mathrm{d}x = \left[-\eta e^{-\lambda x - \lambda y}\right]_0^\infty = \eta e^{-\lambda y}$$

As we can see, the above condition is fulfilled for the calculated values of $p(x)$ and $p(y)$.

$$p(x,y) = p(x) \cdot p(y) = \lambda e^{-\lambda x} \cdot \eta e^{-\lambda y} = \lambda\eta e^{-\lambda x - \eta y} \quad \checkmark .$$

Thus, the variables x and y are statistically independent.

(b)

We define the logarithmic-likelihood function based on the dataset $\mathcal{D} = ((x_1, y_1), ..., (x_N, y_N))$

$$L(\theta) = \ln p(\mathcal{D}, \theta) = \ln \prod_{k=1}^{N} p((x_k, y_k), \theta) = \sum_{k=1}^{N} \ln p((x_k, y_k), \theta) \quad .$$

If we applied this function to the given probability density, under the premise of deriving a maximum likelihood estimator of the parameter λ, we get

$$L(\lambda) = \sum_{k=1}^{N} \ln\left(\lambda\eta e^{-\lambda x_k - \eta y_k}\right) = N \ln(\lambda\eta) - \sum_{k=1}^{N} (\lambda x_k + \eta y_k) \quad .$$

We get the maximum likelihood estimator for λ from the necessary condition $\frac{\partial L(\lambda)}{\partial \lambda} = 0$.

$$0 \stackrel{!}{=} \frac{\partial L(\lambda)}{\partial \lambda} = \frac{N}{\lambda} - \sum_{k=1}^{N} x_k \quad \Leftrightarrow \quad \hat{\lambda} = \sum_{k=1}^{N} \frac{N}{x_k}$$

(c)

Under the constraint $\eta = 1/\lambda$ we get the following logarithmic-likelihood function

$$L(\lambda) = \sum_{k=1}^{N} \ln\Big(e^{-\lambda x_k - \frac{y_k}{\lambda}}\Big) = -\sum_{k=1}^{N} \left(\lambda x_k + \frac{y_k}{\lambda}\right) \quad .$$

From the necessary condition we find

$$0 \overset{!}{=} \frac{\partial L(\lambda)}{\partial \lambda} = -\sum_{k=1}^{N} \left(x_k - \frac{y_k}{\lambda^2}\right) \quad \Leftrightarrow \quad \hat{\lambda} = \sum_{k=1}^{N} \sqrt{\frac{y_k}{x_k}} \quad .$$

(d)

Under the constraint $\eta = 1 - \lambda$ we get the following logarithmic-likelihood function

$$L(\theta) = N \ln\big(\lambda - \lambda^2\big) - \sum_{k=1}^{N} (\lambda x_k + y_k - \lambda y_k) \quad .$$

From the necessary condition we find

$$\begin{aligned} 0 \overset{!}{=} \frac{\partial L(\lambda)}{\partial \lambda} &= \frac{N(1-2\lambda)}{(\lambda - \lambda^2)} - \sum_{k=1}^{N} (x_k - y_k) \\ \Leftrightarrow \qquad &= \underbrace{-\sum_{k=1}^{N} (x_k - y_k)}_{a} \cdot \lambda^2 + \underbrace{\left[\sum_{k=1}^{N} (x_k - y_k) + 2N\right]}_{b} \cdot \lambda - \underbrace{N}_{c} \end{aligned}$$

Using the midnight formula we get the following solutions

$$\hat{\lambda}_{1/2} = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \sum_{k=1}^{N} \frac{(x_k - y_k) + 2N \pm \sqrt{(x_k - y_k)^2 + 4N^2}}{2(x_k - y_k)} \quad .$$

The solutions has to fulfill the condition $\lambda > 0$.

$$\begin{aligned} (x_k - y_k) + 2N &= \sqrt{((x_k - y_k) + 2N)^2} \\ &= \sqrt{(x_k - y_k)^2 + 2N(x_k - y_k) + 4N^2} \\ &> \sqrt{(x_k - y_k)^2 + 4N^2} \end{aligned}$$

With the inequality above we have shown that the two solutions fulfill $\lambda > 0$.

Exercise 2

(a)

We assume that all tosses have been generated independently by the following probability distribution

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

The likelihood function is given by

$$P(\mathcal{D}|\theta) = \prod_{k=1}^{N} P(x_k|\theta) \quad .$$

With the given dataset $\mathcal{D} = (x_1, x_2, ..., x_7) = (\text{head}, \text{head}, \text{tail}, \text{tail}, \text{head}, \text{head}, \text{head})$ we get

$$P(\mathcal{D}|\theta) = \theta^5(1-\theta)^2 \quad .$$

(b)

We apply the logarithmic function to the likelihood function and take its derivative

$$0 \stackrel{!}{=} \frac{\partial}{\partial\theta} \ln\big(\ \theta^5(1-\theta)^2\ \big) = \frac{5-7\theta}{\theta-\theta^2} \quad \Leftrightarrow \quad \hat{\theta} = \frac{5}{7} \quad .$$

From this we can evaluate $P(x_8 = \text{head}, x_9 = \text{head}|\hat{\theta})$.

$$P(x_8 = \text{head}, x_9 = \text{head}|\hat{\theta}) = \hat{\theta}^1(1-\hat{\theta})^0 \cdot \hat{\theta}^1 \cdot (1-\hat{\theta})^0 = \hat{\theta}^2 = \frac{25}{49} \quad .$$

(c)

For a Baysian View, we apply a prior distribution for θ with

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

Our new posterior distribution is

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

The distribution $p(D)$ we get by integrating $P(D|\theta)p(\theta)$ over borders for when θ is equal to one. Inserting the previously computed values, we get

$$\begin{aligned} P(\theta|\mathcal{D}) &= \frac{P(\mathcal{D}|\theta)p(\theta)}{\int_0^1 (D|\theta)p(\theta)d\theta} \\ &= \frac{\theta^5(1-\theta)^2}{\int_0^1 \theta^5(1-\theta)^2 d\theta} \\ &= 168(\theta^5(1-\theta)^2 \quad . \end{aligned}$$

Our new probability for the next two tosses show head is

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

Exercise 3

(a)

The variance of the posterior is given by

$$\sigma_n^2 = \frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2} \quad .$$

To show that the variance of the posterior can be upper-bounded with

$$\sigma_n^2 \leq \min(\frac{\sigma^2}{n}, \sigma_0^2) \quad ,$$

we consider the two cases separate.

<u>Case 1</u>: $\frac{\sigma^2}{n} < \sigma_0^2$

$$\begin{aligned}
\sigma_n^2 &\leq \frac{\sigma^2}{n} \\
\frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2} &\leq \frac{\sigma^2}{n} \\
\frac{\sigma_0^2}{n(\sigma_0^2 + \frac{\sigma^2}{n})} &\leq \frac{1}{n} \\
\sigma_0^2 &\leq \sigma_0^2 + \frac{\sigma^2}{n} \\
0 &\leq \frac{\sigma^2}{n}
\end{aligned}$$

With $\sigma^2 > 0$ and $n \in \mathbb{N}^+$, the statement is $\sigma_n^2 \leq \frac{\sigma^2}{n}$ is always true.

Case 2: $\sigma_0^2 < \frac{\sigma^2}{n}$

$$
\begin{aligned}
\sigma_n^2 &\leq \sigma_0^2 \\
\frac{\sigma_0^2 \sigma^2}{n\sigma_0^2 + \sigma^2} &\leq \sigma_0^2 \\
\frac{\sigma^2}{n\sigma_0^2 + \sigma^2} &\leq 1 \\
\sigma^2 &\leq n\sigma_0^2 + \sigma^2 \\
0 &\leq n\sigma_0^2
\end{aligned}
$$

With $\sigma_0^2 > 0$ and $n \in \mathbb{N}^+$, the statement is $\sigma_n^2 \leq \sigma_0^2$ is always true.

Therefore the variance of the posterior can by upper-bounded with $\sigma_n^2 \leq \min(\frac{\sigma^2}{n}, \sigma_0^2)$.