

MACHINE LEARNING 1: ASSIGNMENT 2

Tom Nick 340528
Niklas Gebauer 340942

Exercise 1

(a) With the definition $P(x_k | \theta)$:

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

We can state the likelihood function $P(\mathcal{D} | \theta)$:

$$\begin{aligned} P(\mathcal{D} | \theta) &= \prod_{k=1}^n P(x_k | \theta) \\ &= \theta^5 \cdot (1 - \theta)^2 \end{aligned}$$

(b) For this task we define:

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

To obtain the maximum likelihood solution $\hat{\theta}$ we can now use the standard procedure to find local minima and maxima in f . So we will at first set the first derivative to zero and solve for θ :

$$\begin{aligned} f'(x) &= 0 \\ \Leftrightarrow (\theta^5 \cdot (1 - \theta)^2)' &= 0 \\ \Leftrightarrow (\theta^7 - 2\theta^6 + \theta^5)' &= 0 \\ \Leftrightarrow 7\theta^6 - 12\theta^5 + 5\theta^4 &= 0 \\ \Leftrightarrow \theta^4(7\theta^2 - 12\theta + 5) &= 0 \\ \Leftrightarrow \theta = 0 \vee 7\theta^2 - 12\theta + 5 &= 0 \end{aligned}$$

So the first possible solution for $\hat{\theta}$ is 0. Now we can use the p-q-formula to calculate the other solutions:

$$\begin{aligned} 7\theta^2 - 12\theta + 5 &= 0 \\ \Leftrightarrow \theta^2 - \frac{12}{7}\theta + \frac{5}{7} &= 0 \\ \Leftrightarrow \frac{12}{14} \pm \sqrt{\left(\frac{12}{14}\right)^2 - \frac{5}{7}} &= \theta_1, \theta_2 \\ \Leftrightarrow \frac{6}{7} \pm \sqrt{\frac{1}{49}} &= \theta_1, \theta_2 \\ \Leftrightarrow 1 = \theta_1 \wedge \frac{5}{7} &= \theta_2 \end{aligned}$$

So we have three possible solutions for $\hat{\theta} \in [0, 1]$:

$$\begin{aligned} \hat{\theta}_0 &= 0 \\ \hat{\theta}_1 &= 1 \\ \hat{\theta}_2 &= \frac{5}{7} \end{aligned}$$

Now we check more derivatives to see which is the local maximum that we are looking for:

$$f''(0) = f^3(0) = f^4(0) = 0 \neq f^5(0) \Rightarrow \text{saddle point at } \hat{\theta}_0$$

$$f''(1) = 2 \Rightarrow \text{minimum at } \hat{\theta}_1$$

$$f''(\frac{5}{7}) = -\frac{1250}{2401} \Rightarrow \text{maximum at } \hat{\theta}_2$$

So the maximum likelihood solution for $\hat{\theta}$ is $\frac{5}{7}$, which does make sense since it is the number of heads compared to the total number of tosses (which is the sample mean, if we assume a random variable X with $X = 1$ if $x_k = \text{head}$ and $X = 0$ otherwise):

$$\hat{\theta} = \frac{\#\{x = \text{head} \mid x \in D\}}{\#(D)} = \frac{5}{7}$$

With this and the fact that each sample x_k is generated independently we can compute $P(x_8 = \text{head}, x_9 = \text{head} \mid \hat{\theta})$:

$$P(x_8 = \text{head}, x_9 = \text{head} \mid \hat{\theta}) = P(x_8 = \text{head} \mid \hat{\theta})P(x_9 = \text{head} \mid \hat{\theta}) = \hat{\theta}^2 = \frac{25}{49} \approx 0.51$$

(c) First we want to compute the posterior distribution $p(\theta \mid \mathcal{D})$. With our prior distribution

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

we get:

$$\begin{aligned} p(\theta \mid \mathcal{D}) &= \frac{p(\mathcal{D} \mid \theta)p(\theta)}{\int p(\mathcal{D} \mid \theta)p(\theta)d\theta} \\ &= \alpha \prod_{k=1}^7 P(x_k \mid \theta)p(\theta) \\ &= \begin{cases} \alpha \cdot \theta^5 \cdot (1 - \theta)^2 & \text{if } 0 \leq \theta \leq 1 \\ 0 & \text{else} \end{cases} \end{aligned}$$

with:

$$\begin{aligned} \alpha &= \frac{1}{\int_0^1 \prod_{k=1}^7 P(x_k \mid \theta)p(\theta)d\theta} \\ &= \frac{1}{\int_0^1 \prod_{k=1}^7 P(x_k \mid \theta)d\theta} \\ &= \frac{1}{\int_0^1 \theta^5 \cdot (1 - \theta)^2 d\theta} \\ &= \frac{1}{\int_0^1 \theta^7 - 2\theta^6 + \theta^5 d\theta} \\ &= \frac{1}{[\frac{\theta^8}{8} - \frac{2\theta^7}{7} + \frac{\theta^6}{6}]_0^1} \\ &= \frac{1}{\frac{28}{168} - \frac{48}{168} + \frac{21}{168}} \\ &= \frac{1}{\frac{1}{168}} = 168 \end{aligned}$$

Now we can evaluate the probability that the next two tosses are head:

$$\begin{aligned}
 \int P(x_8 = \text{head}, x_9 = \text{head} \mid \theta) p(\theta \mid D) d\theta &= \int P(x_8 = \text{head} \mid \theta) P(x_9 = \text{head} \mid \theta) p(\theta \mid D) d\theta \\
 &= \int_0^1 \theta^2 \cdot \alpha \cdot \theta^5 \cdot (1 - \theta)^2 d\theta \\
 &= 168 \int_0^1 \theta^7 (1 - \theta)^2 d\theta \\
 &= 168 \int_0^1 \theta^9 - 2\theta^8 + \theta^7 d\theta \\
 &= 168 \cdot \left[\frac{\theta^{10}}{10} - \frac{2 \cdot \theta^9}{9} + \frac{\theta^8}{8} \right]_0^1 \\
 &= 168 \cdot \frac{1}{360} = \frac{7}{15} \approx 0.4\bar{6}
 \end{aligned}$$