Exercise Sheet 2. Machine learning

(a). $P(D|\theta) = \prod_{m=1}^{7} P(x_m|\theta)$

$= \theta \cdot \theta \cdot (1-\theta) \cdot (1-\theta) \cdot \theta \cdot \theta \cdot \theta$

$= \theta^5 \cdot (1-\theta)^2$. //.

(b) by $P(D|\theta) = \theta^5 \cdot (1-\theta)^2$:

$\hat{\theta} := \arg\max_\theta P(D|\theta)$

$= \arg\max_\theta (\theta^5 (1-\theta)^2)$.

let $L(\theta) = \theta^5 \cdot (1-\theta)^2$:

$\nabla_\theta L(\theta) = \frac{d(\theta^7 - 2\theta^6 - \theta^5)}{d\theta} = 7\theta^6 - 12\theta^5 - 5\theta^4 = \theta^4 (7\theta - 5)(\theta - 1)$

let $\nabla_\theta L(\theta) = 0$:

$\theta = 0$ or 1 or $\frac{5}{7}$

since $0 < \theta < 1$ is reasonable: $\theta = \frac{5}{7}$.

so: $P(x_8 = \text{head}, x_9 = \text{head} | \hat{\theta}) = \theta^2 = \frac{25}{49}$. //

(c). $P(\theta|D) = \frac{P(D|\theta) \cdot P(\theta)}{\int P(D|\theta) \cdot P(\theta) \cdot d\theta}$

by: $\int P(D|\theta) \cdot P(\theta) d\theta = \int_0^1 \theta^5 (1-\theta)^2 d\theta = \frac{1}{168}$

so: $P(\theta|D) = \begin{cases} \theta^5 (1-\theta)^2 \cdot 168 & \text{if } 0 \le \theta \le 1. \\ 0 & \text{else}. \end{cases}$

so: $\int P(x_8 = \text{head}, x_9 = \text{head} | \theta) \cdot P(\theta|D) \, d\theta$

$= \int_0^1 \theta^2 \cdot \theta^5 (1-\theta)^2 \cdot 168 \, d\theta$

$= 168 \left(\frac{1}{10}\theta^{10} - \frac{2}{9}\theta^9 + \frac{1}{8}\theta^8 \right) \Big|_0^1$

$= \frac{168}{360}$

$= \frac{7}{15}$

2. (a) by $\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$:

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

①. $n\sigma_0^2 \le n\sigma_0^2 + \sigma^2$.

so: $\frac{\sigma_0^2}{n\sigma_0^2 + \sigma^2} \le \frac{1}{n}$.

$\frac{\sigma^2 \cdot \sigma_0^2}{n\sigma_0^2 + \sigma^2} \le \frac{\sigma^2}{n}$. ie: $\sigma_n^2 \le \frac{\sigma^2}{n}$

② $\sigma^2 \le n\sigma_0^2 + \sigma^2$.

so: $\frac{\sigma^2}{n\sigma_0^2 + \sigma^2} \le 1$.

$\frac{\sigma^2 \cdot \sigma_0^2}{n\sigma_0^2 + \sigma^2} \le \sigma_0^2$. ie: $\sigma_n^2 \le \sigma_0^2$

in conclusion: $\sigma_n^2 \le \min(\frac{\sigma^2}{n}, \sigma_0^2)$. //

(b) by $\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$ and $\frac{u_n}{\sigma_n^2} = \frac{n}{\sigma^2}\hat{u}_n + \frac{u_0}{\sigma_0^2}$:

$$u_n = \frac{n \cdot \sigma_n^2}{\sigma^2} \cdot \hat{u} + \frac{u_0}{\sigma_0^2} \cdot \sigma_n^2$$

$$= \frac{n\sigma_0^2}{n\sigma_0^2 + \sigma^2} \cdot \hat{u} + \frac{u_0 \cdot \sigma^2}{n\sigma_0^2 + \sigma^2}$$

①. if $\hat{u} \le u_0$, then: $n\hat{u}\sigma_0^2 \le n u_0 \sigma_0^2$.

$$u_n = \frac{n\sigma_0^2 \hat{u} + u_0 \sigma^2}{n\sigma_0^2 + \sigma^2} \le \frac{n u_0 \sigma_0^2 + u_0 \sigma^2}{n\sigma_0^2 + \sigma^2} = u_0$$

$u_0 \sigma^2 \ge \hat{u}\sigma^2$.

$$u_n = \frac{n\sigma_0^2 \hat{u} + u_0 \sigma^2}{n\sigma_0^2 + \sigma^2} \ge \frac{n\sigma_0^2 \hat{u} + \hat{u}\sigma^2}{n\sigma_0^2 + \sigma^2} = \hat{u}$$

② if $\hat{u} \ge u_0$, then:

$$u_n \le \frac{n\sigma_0^2 \hat{u} + \hat{u}\sigma^2}{n\sigma_0^2 + \sigma^2} = \hat{u}$$

$$u_n \ge \frac{n\sigma_0^2 \cdot u_0 + u_0 \sigma^2}{n\sigma_0^2 + \sigma^2} = u_0$$

in conclusion,

$\min(\hat{u}_n, u_0) \le u_n \le \max(\hat{u}_n, u_0)$ //