Machine Learning 1 WS18/19

Submission for Exercise Sheet 2

Hendrik Makait 384968
Michael Hoppe 362514
Wai Tang Victor Chan 406094
Rudi Poepsel Lemaitre 373017
Jonas Piotrowski 399334
Aki Saksala 399293

ML 1 Exercise 2

1a) for independence: $p(x,y) = p(x)\,p(y) \quad \forall\, x,y \in \mathbb{R}^2$

$$p(x) = \int_0^\infty p(x,y)\,dy = \int_0^\infty \lambda\eta e^{-\lambda x - \eta y}\,dy$$

$$= \lambda\eta e^{-\lambda x} \int e^{-\eta y}\,dy = \lambda\eta e^{-\lambda x}\left[\frac{-1}{\eta} e^{-\eta y}\right]_0^\infty = \lambda\eta e^{-\lambda x}$$

$$p(y) = \int_0^\infty p(x,y)\,dx = \int_0^\infty \lambda\eta e^{-\lambda x - \eta y}\,dx = \lambda\eta e^{-\eta y}\int_0^\infty e^{-\lambda x}\,dx$$

$$= \lambda\eta e^{-\eta y}\left[\frac{-1}{\lambda} e^{-\lambda x}\right]_0^\infty = \eta e^{-\eta y}$$

$$\Rightarrow p(x,y) = p(x)\,p(y) = \lambda\eta e^{-\lambda x - \eta y} \quad \square$$

1b) Likelihood-funktion: $p(D|\lambda,\eta) = \prod_1^N p(x_i, y_i \mid \lambda) = \prod_1^N \lambda\eta e^{-\lambda x_i - \eta y_i}$

$$\ell(\lambda) = \log(p(D|\lambda,\eta)) =$$

$$= \sum_{i=1}^N \left(\log(\lambda) + \log(\eta) - \lambda x_i - \eta y_i\right)$$

$$= \cancel{N(\log(\lambda) + \log(\eta))}$$

$$= N(\log(\lambda\eta) - \sum_1^N \lambda x_i + \eta y_i$$

$$\arg\max_\lambda \ell(\lambda) \Rightarrow \frac{\ell(\lambda)}{d\lambda} = \frac{N}{\lambda} - \sum_{i=1}^N x_i \overset{!}{=} 0$$

$$\Leftrightarrow \frac{N}{\lambda} = \sum_1^N x_i$$

$$\Leftrightarrow \hat{\lambda} = \frac{\sum_1^N (x_i)}{N}$$

A

1)c $p(D \mid \lambda, \eta) = \prod_1^N p(x_i\, y_i \mid \lambda) = \prod_1^N \lambda\eta\, e^{-\lambda x_i - \eta y_i}$

$$\Rightarrow \eta = \frac{1}{\lambda} \quad \Rightarrow p(D \mid \lambda) = \prod_1^N e^{-\lambda x_i - \frac{1}{\lambda} y_i}$$

$$\ell(\lambda) = \log p(D \mid \lambda) = \sum_1^N -\lambda x_i - \frac{1}{\lambda} y_i$$

$$\frac{d\ell(\lambda)}{d\lambda} = -\sum_1^N x_i + \frac{1}{\lambda^2}\sum_1^N y_i \overset{!}{=} 0$$

$$\lambda^2 = \frac{\sum_1^N y_i}{\sum_1^N x_i}$$

$$\lambda = \sqrt{\frac{\sum_1^N y_i}{\sum_1^N x_i}}$$

1d) $p(D \mid \lambda, \eta) = \prod_1^N p(x_i, y_i \mid \lambda) = \prod_1^N \lambda . \eta\, e^{-\lambda x_i - \eta y_i}$

$$\Rightarrow \eta = 1 - \lambda \qquad \Rightarrow p(D \mid \lambda) = \prod_1^n \lambda(1-\lambda) e^{-\lambda x_i - y_i + \lambda y_i}$$

$$\ell(\lambda) = \sum_1^N \log(\lambda - \lambda^2) - \lambda x_i - y_i + \lambda y_i$$

$$= \log(\lambda - \lambda^2) - N y_i + \sum_1^N + \lambda y_i - \lambda x_i$$

$$\frac{d\ell(\lambda)}{d\lambda} = \frac{1}{\lambda} - \frac{1}{1-\lambda} + \sum_1^N y_i - x_i \overset{!}{=} 0$$

$$\sum_1^N y_i - x_i = \frac{1}{1-\lambda} - \frac{1}{\lambda}$$

$$\sum_1^N y_i - x_i = \frac{\lambda - 1 + \lambda}{\lambda(1-\lambda)}$$

$$\lambda \sum_1^N y_i - x_i - \lambda^2 \sum_1^N y_i - x_i = 2\lambda - 1$$

$$\lambda^2 \sum_1^N (y_i - x_i) + \lambda\left(2 - \sum_1^N (y_i - x_i)\right) + 1 = 0$$

$$\lambda = \frac{\sum_1^N (y_i - x_i) - 2 \pm \sqrt{4 - 4\Sigma + \Sigma^2 + 4 \cdot \Sigma \cdot 1}}{2\sum_1^N (y_i - x_i)}$$

$$\lambda = \frac{\sum_1^N (y_i - x_i) - 2 \pm \sqrt{4 + \left(\sum (y_i - x_i)\right)^2}}{2\sum_1^N (y_i - x_i)}$$

ML Exercise 2

(2) a)

$$P(D|\theta) = \prod_{n=1}^{N} p(x_n|\theta)$$

$$= p(\text{head}|\theta)^5 \, p(\text{tail}|\theta)^2$$

$$= \theta^5 (1-\theta)^2$$

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

$$P(D|\theta) = \theta^7 - 2\theta^6 + \theta^5$$

b) $\hat{\theta}$

$$0 = 7\theta^6 - 12\theta^5 + 5\theta^4$$

$$\hat{\theta} = \frac{5}{7}$$

$$P(x_8 = \text{head}, x_9 = \text{head} | \hat{\theta}) = p(\text{head}|\hat{\theta})^2$$

$$= \frac{5}{7}^2$$

$$= \frac{25}{49} \approx \underline{0{,}51}$$

c)

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

$$p(\theta|D) = \frac{p(D|\theta)\,p(\theta)}{p(D)} \qquad p(\theta|D) = \frac{p(D|\theta)\,p(\theta)}{\int p(D|\theta)p(\theta)\,d\theta}$$

$$p(\theta|D) = \frac{\prod_{n=1}^{N} p(x_n|\theta) \cdot 1}{\int \prod_{n=1}^{N} p(x_n|\theta) \cdot 1 \, d\theta} = \frac{p(\text{head}|\theta)^5 \, p(\text{tail}|\theta)^2}{\int p(\text{head}|\theta)^5 \, p(\text{tail}|\theta)^2 \, d\theta}$$

$$= \frac{\theta^5 (1-\theta)^2}{\int_0^1 \theta^5 (1-\theta)^2 \, d\theta}$$

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

$$= \int_0^1 \theta^7 - 2\theta^6 + \theta^5$$

$$= \frac{1}{8}\cdot 1^8 - \frac{2}{7} 1^7 + \frac{1}{6} 1^6$$

$$= \frac{1}{8} - \frac{2}{7} + \frac{1}{6}$$

$$= \frac{1}{168} \approx 0{,}0059524$$

$$p(\theta|D) = \frac{\theta^5(1-\theta)^2}{\frac{1}{168}} = \underline{\underline{168\,\theta^5(1-\theta)^2}}$$

$$\boxed{p(x|D) = \int p(x|\theta)\, p(\theta|D)\, d\theta}$$

$$\int P(x_8 = head, x_9 = head\,|\,\theta)\, p(\theta|D)\, d\theta$$

$$= \int p(head|\theta)^2 \cdot 168\,\theta^5(1-\theta)^2\, d\theta$$

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

$$= \int 168\cdot(\theta^7 - 2\theta^6 + \theta^5)\cdot\theta^2\, d\theta$$

$$= \int (168\,\theta^7 - 336\,\theta^6 + 168\,\theta^5)\,\theta^2\, d\theta$$

$$= \int_0^1 168\,\theta^9 - 336\,\theta^8 + 168\,\theta^7\, d\theta$$

$$= \frac{168}{10} 1^{10} - \frac{336}{9} 1^9 + \frac{168}{8} 1^8$$

$$= \frac{84}{5} - \frac{112}{3} + 21$$

$$\approx \underline{\underline{0{,}4667}}$$

Exercise 3:

a) Show: $\sigma_n^2 \leq \min\left(\frac{\sigma^2}{n}, \sigma_0^2\right)$

It is given: $\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$, $\sigma^2, \sigma_0^2 > 0$

(i) $\overset{\sigma^2>0}{\Longrightarrow} \frac{1}{\sigma_n^2} \geq \frac{n}{\sigma^2} \overset{\sigma_n^2, \sigma^2, n > 0}{\iff} \frac{\sigma^2}{n} \geq \sigma_n^2$

as well as

(ii) $\overset{\sigma_0^2>0}{\Longrightarrow} \frac{1}{\sigma_n^2} \geq \frac{1}{\sigma_0^2} \overset{\sigma_n^2, \sigma_0^2 > 0}{\iff} \sigma_0^2 \geq \sigma_n^2$

$\overset{(i),(ii)}{\Longrightarrow} \sigma_0^2 \geq \sigma_n^2$ and $\sigma^2 \geq \sigma_n^2 \Rightarrow \min(\sigma^2, \sigma_0^2) \geq \sigma_n^2$ q.e.d.

b) Show: $\min(\hat{\mu}_n, \mu_0) \leq \mu_n \leq \max(\hat{\mu}_n, \mu_0)$

with given (1): $\frac{1}{\sigma_n^2} = \frac{n}{\sigma^2} + \frac{1}{\sigma_0^2}$, (2): $\frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2}$

Case 1: $\hat{\mu}_n \geq \mu_0 \iff \hat{\mu}_n = \max(\hat{\mu}_n, \mu_0)$, $\mu_0 = \min(\hat{\mu}_n, \mu_0)$

~~It is given: (1)~~ $\cancel{\frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma}}$

(i) (2): $\frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \geq \frac{n}{\sigma^2}\mu_0 + \frac{1}{\sigma_0^2}\mu_0 \overset{(1)}{=} \frac{1}{\sigma_n^2}\mu_0$

$\iff \frac{\mu_n}{\sigma_n^2} \geq \frac{\mu_0}{\sigma_n^2} \iff \mu_n \geq \mu_0 = \min(\hat{\mu}_n, \mu_0)$

(ii) (2) $\frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \leq \frac{n}{\sigma^2}\hat{\mu}_n + \frac{1}{\sigma_0^2}\hat{\mu}_n \overset{(1)}{=} \frac{1}{\sigma_n^2}\hat{\mu}_n$

$\iff \frac{\mu_n}{\sigma_n^2} \leq \frac{\hat{\mu}_n}{\sigma_n^2} \iff \mu_n \leq \hat{\mu}_n = \max(\hat{\mu}_n, \mu_0)$

Case 2: $\hat{\mu}_n < \mu_0 \Leftrightarrow \hat{\mu}_n = \min(\hat{\mu}_n, \mu_0), \mu_0 = \max(\hat{\mu}_n, \mu_0)$

(iii) (2) $$\frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \geq \frac{n}{\sigma^2}\hat{\mu}_n + \frac{1}{\sigma_0^2}\hat{\mu}_n \overset{(1)}{=} \frac{1}{\sigma_n^2}\hat{\mu}_n$$

$$\Leftrightarrow \frac{\mu_n}{\sigma_n^2} \geq \frac{\hat{\mu}_n}{\sigma_n^2} \Leftrightarrow \mu_n \geq \hat{\mu}_n = \min(\hat{\mu}_n, \mu_0)$$

(iv) (2) $$\frac{\mu_n}{\sigma_n^2} = \frac{n}{\sigma^2}\hat{\mu}_n + \frac{\mu_0}{\sigma_0^2} \leq \frac{n}{\sigma_n^2}\mu_0 + \frac{\mu}{\sigma_0^2} \overset{(1)}{=} \frac{1}{\sigma_n^2}\mu_0$$

$$\Leftrightarrow \frac{\mu_n}{\sigma_n^2} \leq \frac{\mu_0}{\sigma_n^2} \Leftrightarrow \mu_n \leq \mu_0 = \max(\hat{\mu}_n, \mu_0)$$

From (i) to (iv) it follows that

$$\min(\hat{\mu}_n, \mu_0) \leq \mu_n \leq \max(\hat{\mu}_n, \mu_0)$$ q.e.d.