G. Montavon
Institute of Computer Science
Department of Mathematics and Computer Science
Freie Universität Berlin

Exercise Sheet 9 (theory part)

Exercise 1: Maximum-Likelihood Estimation (5 + 5 + 5 + 5 P)

We consider the problem of estimating using the maximum-likelihood approach the parameters $\lambda, \eta > 0$ of the probability distribution:

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

supported on $\mathbb{R}_+^2$. We consider a dataset $\mathcal{D} = ((x_1, y_1), \ldots, (x_N, y_N))$ composed of N independent draws from this distribution.

(a) *Show* that x and y are independent.

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

(b) *Derive* a maximum likelihood estimator of the parameter λ based on $\mathcal{D}$.

$$\begin{aligned} J(\theta, \lambda) &= \log p(\mathcal{D}|\theta) \\ &= \sum_{k=1}^{N} \log p(x_k, y_k|\lambda, \eta) \\ &= \sum_{k=1}^{N} \big[\log \lambda - \lambda x_k + \log \eta - \eta y_k\big] \\ &= N \cdot \big(\log \lambda - \lambda \bar{x} + \log \eta - \eta \bar{y}\big) \qquad \text{concave with } \lambda \end{aligned}$$

$$\frac{\partial J}{\partial \lambda} = N \cdot \Big(\frac{1}{\lambda} - \bar{x}\Big) = 0 \quad \Rightarrow \quad \boxed{\lambda = \frac{1}{\bar{x}}}$$

(c) *Derive* a maximum likelihood estimator of the parameter λ based on $\mathcal{D}$ subject to $\eta = 1/\lambda$.

$$\begin{aligned} J &= N\big(\log \lambda - \lambda \bar{x} + \log \eta - \eta \bar{y}\big) \\ &= N\big(-\lambda \bar{x} - \frac{1}{\lambda}\bar{y}\big) \qquad \text{concave for } \lambda > 0 \end{aligned}$$

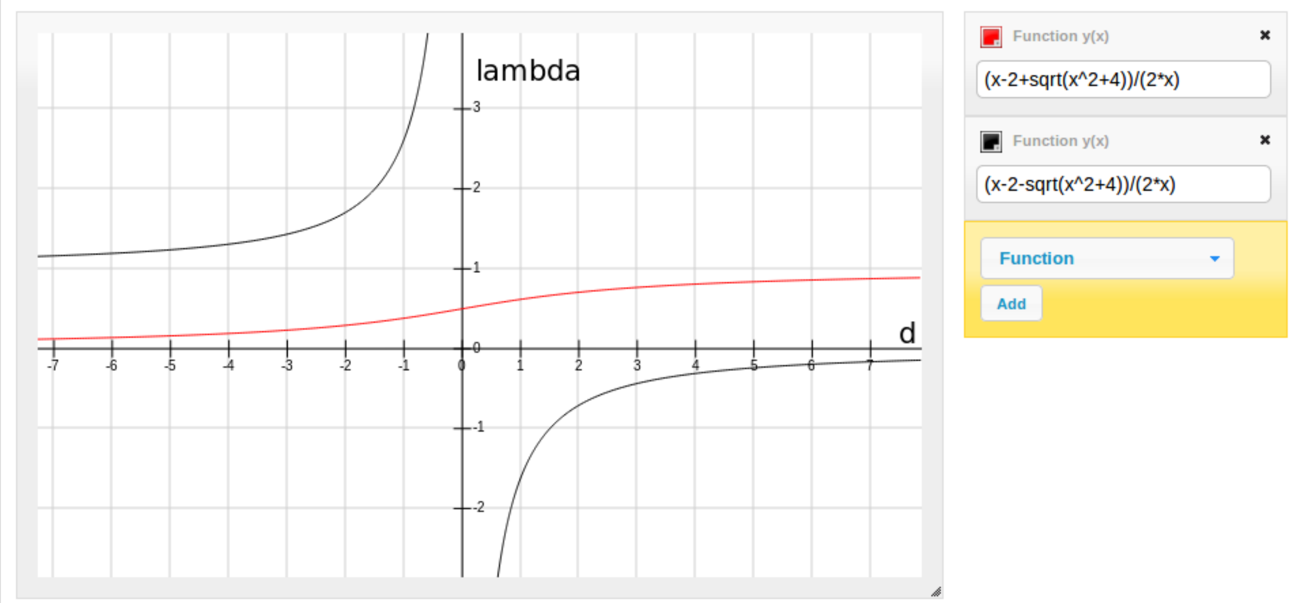
$$\frac{\partial J}{\partial \lambda} = N \cdot \Big(-\bar{x} + \frac{\bar{y}}{\lambda^2}\Big) = 0 \quad \Rightarrow \quad \lambda = \pm\sqrt{\frac{\bar{y}}{\bar{x}}} \quad \Rightarrow \quad \boxed{\lambda = \sqrt{\frac{\bar{y}}{\bar{x}}}}$$

(d) *Derive* a maximum likelihood estimator of the parameter λ based on $\mathcal{D}$ subject to $\eta = 1 - \lambda$.

$$
\begin{aligned}
J &= N\big(\log\lambda - \lambda\bar{x} + \log(1-\lambda) - (1-\lambda)\bar{y}\big)\\
&= N\big(\log(\lambda-\lambda^2) + \lambda\underbrace{(\bar{y}-\bar{x})}_{\bar{d}}\big) + \text{cst.}
\end{aligned}
$$

$$
\frac{\partial J}{\partial \lambda} = N \cdot \Big(\frac{1-2\lambda}{\lambda-\lambda^2} + \bar{d}\Big) = 0
$$

$$
\lambda = \frac{(\bar{d}-2) \pm \sqrt{\bar{d}^2+4}}{2\bar{d}}
$$



Exercise 2: Maximum Likelihood vs. Bayes ($5+10+15$ P)

An unfair coin is tossed seven times and the event (head or tail) is recorded at each iteration. The observed sequence of events is

$$
\mathcal{D} = (x_1, x_2, \dots, x_7) = (\text{head}, \text{head}, \text{tail}, \text{tail}, \text{head}, \text{head}, \text{head}).
$$

We assume that all tosses $x_1, x_2, \dots$ 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.

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

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

$$
p(\mathcal{D}|\theta) = \prod_{k=1}^{7} p(x_k|\theta) = \theta \cdot \theta \cdot (1-\theta) \cdot (1-\theta) \cdot \theta \cdot \theta \cdot \theta = \theta^5 \cdot (1-\theta)^2
$$

$$
\log p(\mathcal{D}|\theta) = 5\log\theta + 2\log(1-\theta) \qquad \text{concave}
$$

$$
\frac{\partial}{\partial\theta}\log p(\mathcal{D}|\theta) = \frac{5}{\theta} - \frac{2}{1-\theta} = 0 \quad \Rightarrow \quad \hat{\theta} = \frac{5}{7}
$$

$$
P(x_8 = \text{head}\,,\, x_9 = \text{head} \mid \hat{\theta}) = P(x_8 = \text{head} \mid \hat{\theta}) \cdot P(x_9 = \text{head} \mid \hat{\theta}) = \hat{\theta} \cdot \hat{\theta} = \frac{5}{7} \cdot \frac{5}{7} = \frac{25}{49}
$$

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

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

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

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

$$p(\theta|\mathcal{D}) = \frac{p(\mathcal{D}|\theta)p(\theta)}{\int p(\mathcal{D}|\theta)p(\theta)d\theta} = \frac{\theta^5 \cdot (1-\theta)^2}{\int_0^1 \theta^5 \cdot (1-\theta)^2 \cdot 1} = 168 \cdot \theta^5 \cdot (1-\theta)^2 \qquad (0 \leq \theta \leq 1)$$

$$\int P(x_8 = \text{head}\,,\, x_9 = \text{head} \mid \theta)\, p(\theta|\mathcal{D})\, d\theta = \int_0^1 \theta^2 \cdot 168 \cdot \theta^5 \cdot (1-\theta)^2 \cdot 1 \cdot d\theta = \frac{168}{360} = \frac{7}{15}$$