Winter term 2020/21
L. Rendsburg / U. von Luxburg

Assignment 11

Mathematics for Machine Learning

Submission due on **08.02.21, 8:00**

Justify all your claims.

Exercise 1 (**Estimation**, 2+1+2 points)**.** Let $X \sim \mathrm{Pois}(-1/2 \log \theta)$ be a Poisson-distributed random variable with $\theta \in \Theta = (0,1)$. We now want to estimate θ based on one sample of X.

a) Consider the estimator $U = (-1)^X$. Prove that U is the only unbiased estimator for θ.
Hint: Use the equality $\exp(x) = \sum_{k=0}^{\infty} \frac{x^k}{k!}$. Additionally, use the fact that

$$\left(\sum_{k=0}^{\infty} a_k \frac{x^k}{k!} = \sum_{k=0}^{\infty} b_k \frac{x^k}{k!} \quad \forall x \in (0, \infty) \right) \quad \Rightarrow \quad a_k = b_k \quad \forall k \in \mathbb{N}_0 \, .$$

b) What is the $\mathrm{MSE}(U, \theta)$?

c) Now consider another estimator $V = \mathbb{1}_{2\mathbb{N}_0}(X)$. Is V unbiased? Prove that $\mathrm{MSE}(V, \theta) < \mathrm{MSE}(U, \theta)$ for all $\theta \in \Theta$.
Hint: Show that $|V(x) - \theta| \leq |U(x) - \theta|$ for all $x \in \mathbb{N}_0$, where the inequality is strict ($<$) if $x \in 2\mathbb{N}_0 + 1$.

Exercise 2 (**Sufficiency and exponential families**, 3+2+3+2 points)**.** Consider a parametric family $\mathcal{F} = \{f_\theta \mid \theta \in \Theta\}$ of densities on $\mathbb{R}^d$. A test statistic $T \colon \mathbb{R}^d \to \mathbb{R}^l$ is called *sufficient*, if the distribution of X given T does not depend on θ, that is, $f_\theta(x | T(X) = T(x))$ does not depend on θ for all $x \in \mathbb{R}^d$.

a) The factorization theorem states that T is sufficient, if and only if there exist functions $h, (g_\theta)_{\theta \in \Theta}$ such that the densities can be decomposed as

$$f_\theta(x) = h(x) \cdot g_\theta(T(x)) \quad \forall \theta \in \Theta \, . \tag{1}$$

Prove the backward implication of the factorization theorem for the special case of discrete densities. That is, assume that there exists a countable subset $\mathcal{X} \subseteq \mathbb{R}^d$ with $P_\theta(X \in \mathcal{X}) = \sum_{x \in \mathcal{X}} f_\theta(x) = 1$ for every $\theta \in \Theta$. Then prove that Eq. (1) implies that T is sufficient.

b) Consider $n \in \mathbb{N}$ i.i.d. samples $X_1, \ldots, X_n$ from exponential distributions $\mathcal{F}_{exp} = \{f_\lambda(x) = \lambda \exp(-\lambda x) \mid \lambda > 0\}$ on $(0, \infty)$. Use the factorization theorem to show that $S(x_1, \ldots, x_n) = \sum_{i=1}^n x_i$ is sufficient.

c) $\mathcal{F}$ is called a *k-dimensional exponential family*, if there exist measurable functions $d, c_1, \ldots, c_k \colon \Theta \to \mathbb{R}$ and $h, T_1, \ldots, T_k \colon \mathbb{R}^d \to \mathbb{R}$ such that

$$f_\theta(x) = h(x) \cdot \exp\left(\sum_{j=1}^{k} c_j(\theta) T_j(x) + d(\theta) \right) \quad \forall \theta \in \Theta, x \in \mathbb{R}^d \, .$$

Consider $n \in \mathbb{N}$ i.i.d. samples $X_1, \ldots, X_n$. Use the factorization theorem on $\tilde{\mathcal{F}} = \left\{ \tilde{f}_\theta(x_1, \ldots, x_n) = \prod_{i=1}^n f_\theta(x_i) \mid \theta \in \Theta \right\}$ to prove that the statistic

$$T \colon \left(\mathbb{R}^d\right)^n \to \mathbb{R}^k,$$
$$x = (x_1, \ldots, x_n) \mapsto \left(\sum_{i=1}^n T_1(x_i), \ldots, \sum_{i=1}^n T_k(x_i) \right)$$

is sufficient.

d) Show that $\mathcal{F}_{\mathcal{N}} = \{f_{(\mu,\sigma^2)} \mid \mu \in \mathbb{R}, \sigma^2 > 0\}$ is a 2-dimensional exponential family, where $f_{(\mu,\sigma^2)}$ is the density of a normal distribution $\mathcal{N}(\mu, \sigma^2)$. Use part c) to find a sufficient statistic for $\theta = (\mu, \sigma^2)$ based on n i.i.d. samples $X_1, \dots, X_n$.

Exercise 3 (**Maximum likelihood estimation**, 2+3 points)**.**

a) Consider i.i.d. samples $X_1, \dots, X_n$ from an exponential distribution $\mathrm{Exp}(\lambda)$ on $(0, \infty)$ with $\lambda > 0$. Find the maximum likelihood estimator $\hat{\lambda} = \hat{\lambda}(X_1, \dots, X_n)$.

b) Now assume the setting in a), but we only observe the censored random variables $Y_k = \min\{X_k, c\}$ for $k = 1, \dots, n$ and some $c > 0$. Find the maximum likelihood estimator $\tilde{\lambda} = \tilde{\lambda}(Y_1, \dots, Y_n)$ based on these censored random variables.
Hint: Consider a mixed distribution $P_\lambda = \nu_\lambda^{(\mathrm{cont})} + \nu_\lambda^{(\mathrm{sing})}$ on $(0, \infty)$ with continuous part $\nu_\lambda^{(\mathrm{cont})}$, singular part $\nu_\lambda^{(\mathrm{sing})}$, and corresponding densities $f_\lambda^{(cont)}, f_\lambda^{(sing)}$ (compare Decomposition by Lebesgue). Define $\mathcal{Y}^{(\mathrm{sing})} = \{y \in (0, \infty) \mid f_\lambda^{(\mathrm{sing})}(y) > 0\}$. The likelihood function for a sample $y_1, \dots, y_n$ is then given by

$$L(\lambda; y_1, \dots, y_n) = \prod_{\substack{i=1 \\ y_i \notin \mathcal{Y}^{(\mathrm{sing})}}}^{n} f_\lambda^{(cont)}(y_i) \prod_{\substack{i=1 \\ y_i \in \mathcal{Y}^{(\mathrm{sing})}}}^{n} f_\lambda^{(sing)}(y_i)\,.$$