STATISTICAL MACHINE LEARNING (WS2022)
SEMINAR 6

Assignment 1. Consider the family of Bernoulli distributions for a binary random variable $x = 0, 1$ given by

$$p(x; \beta) = \beta^x (1 - \beta)^{1-x},$$

where $\beta \in (0, 1)$ is the parameter. Prove that it is an exponential family. Give the sufficient statistics (***Hint:*** it is one dimensional) and express the natural parameter as a function of β. Give the cumulant function as a function of the natural parameter.

Assignment 2. Prove that the family of univariate normal distributions $\mathcal{N}(\mu, \sigma)$ with density

$$p(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

is an exponential family

$$p_\eta(x) = \exp\big[\langle \phi(x), \eta \rangle - A(\eta)\big]$$

with sufficient statistics $\phi(x) = \begin{bmatrix} x \\ x^2 \end{bmatrix}$. Deduce a formula expressing the natural parameter vector η in terms of μ and σ.

Assignment 3. The Kullback-Leibler divergence for probability densities $p(x)$ and $q(x)$ is defined by

$$D_{KL}(p \parallel q) = \int_{-\infty}^{\infty} p(x) \log \frac{p(x)}{q(x)}\, dx.$$

Compute the KL-divergence for two univariate normal distributions $\mathcal{N}(\mu, \sigma)$ and $\mathcal{N}(\tilde{\mu}, \tilde{\sigma})$.

Assignment 4. The probability density function of a Laplace distribution (aka double exponential distribution) with location parameter μ and scale b is given by

$$p(x \mid \mu, b) = \frac{1}{2b} \exp\Big(-\frac{|x - \mu|}{b}\Big).$$

Find the maximum likelihood estimates of the location parameter and the scale parameter given an i.i.d. sample $\mathcal{T}^m = \{x_i \in \mathbb{R} \mid i = 1, \dots, m\}$.

Assignment 5. Consider the cumulant function $A(\eta)$ of an exponential family

$$p_\eta(x) = \exp\big[\phi(x)\eta - A(\eta)\big]$$

for a discrete random variable $x \in \mathcal{X}$. It is defined by

$$A(\eta) = \log \sum_{x \in \mathcal{X}} \exp\big[\phi(x)\eta\big].$$

Notice that we consider for simplicity that $\phi(x)$ and η are scalars.

a) Prove that its first derivative is given by

$$\frac{d}{d\eta}A(\eta) = \mathbb{E}_{x\sim p_\eta}[\phi(x)].$$

b) Prove that its second derivative is non-negative and conclude that $A(\eta)$ is a convex function.

Assignment 6. Consider the family of univariate normal distributions $\mathcal{N}(\mu, 1)$ with density

$$p(x;\mu) = \frac{1}{\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2}}.$$

Compute its Fisher information $I(\mu)$. Suppose you want to estimate the mean μ from a sample $\mathcal{T}^m$. What sample size m is needed (asymptotically) to ensure that the estimated mean will be in the ϵ interval around the true mean with probability 99%?

Assignment 7. * Prove that under the regularity assumptions given in the lecture, the Fisher information of a parametric distribution family $p_\theta(x)$ can be equivalently written as

$$I(\theta) = \mathbb{V}_\theta\Big[\frac{d}{d\theta}\log p_\theta(x)\Big] \quad \text{and} \quad I(\theta) = -\mathbb{E}_\theta\Big[\frac{d^2}{d\theta^2}\log p_\theta(x)\Big].$$

a) Prove that $\mathbb{E}_\theta\Big[\frac{d}{d\theta}\log p_\theta(x)\Big] = 0$ and conclude the first equivalent definition above.

b) Use integration by parts and the logarithmic "trick" $\frac{d}{d\theta}p_\theta(x) = p_\theta(x)\frac{d}{d\theta}\log p_\theta(x)$ to prove the second equivalent definition above.