STATISTICAL MACHINE LEARNING (WS2024)
GENERATIVE LEARNING

Assignment 1. 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 2. 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 $\mathcal{T}^m = \{x_i \in \mathbb{R} \mid i = 1, \ldots, m\}$.

Assignment 3. 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 4. 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 5. Prove the equality $\mathbb{E}_\theta\left[\frac{d}{d\theta}\log p_\theta(x)\right] = 0$ and conclude that the Fisher information is the variance $I(\theta) = \mathbb{V}_\theta\left[\frac{d}{d\theta}\log p_\theta(x)\right]$ of the random variable $\frac{d}{d\theta}\log p_\theta(x)$.