STATISTICAL MACHINE LEARNING (WS2019)
SEMINAR 5

Assignment 1. Consider the following parameter estimation task. You are given i.i.d. training data $\mathcal{T}^m = \{x_i \in \mathbb{R} \mid i = 1, 2, \ldots, m\}$ generated from the normal distribution

$$p_{\mu_0}(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-\mu_0)^2}{2}}$$

and the task is to estimate its unknown mean μ_0.

a) Prove that the expected log-likelihood

$$L(\mu) = \int_{-\infty}^{\infty} p_{\mu_0}(x) \log p_\mu(x)\, dx$$

has a unique global maximum at μ_0.

b) Show that the maximum likelihood estimator

$$e_{ML}(\mathcal{T}^m) = \arg\max_\mu \frac{1}{m} \sum_{i=1}^m \log p_\mu(x).$$

gives the arithmetic mean of the training data, i.e.

$$\mu^* = e_{ML}(\mathcal{T}^m) = \frac{1}{m} \sum_{i=1}^m x_i,$$

Prove that this estimator is unbiased.

c) Compute the variance of the maximum likelihood estimator, i.e. $\mathbb{V}_{\mu_0}\big[e_{ML}(\mathcal{T}^m))\big]$. How does it depend on μ_0 and m?

Assignment 2. The probability density function of a Laplace distribution (aka double exponential distribution) with location parameter μ and scale b is given by

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

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 binary logistic regression model

$$p(y \mid x) = \frac{e^{y\langle w,x\rangle}}{2\cosh\langle w,x\rangle},$$

where $x \in \mathbb{R}^n$ is a feature vector and $y = \pm 1$ is the object class. The model is parametrised by the vector $w \in \mathbb{R}^n$. Given an i.i.d. training set $\mathcal{T}^m = \{(x^j, y^j) \mid$

$j = 1, \ldots, m\}$, we want to estimate the unknown parameter vector w of the model by maximising the conditional log-likelihood

$$\begin{aligned} L(w, \mathcal{T}^m) =& \frac{1}{m} \sum_{j=1}^{m} \log(p(y \mid x)) = \\ =& \frac{1}{m} \sum_{j=1}^{m} \Big[y^j \left\langle w, x^j \right\rangle - \log\cosh \left\langle w, x^j \right\rangle - \log 2 \Big] \to \max_w \end{aligned}$$

Prove that the objective function is concave in w by computing its second derivative (matrix) and showing that it is negative semidefinite.

Assignment 4. Consider the Kullback-Leibler divergence for probability densities $p(x)$ and $q(x)$ defined by

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

Let us assume for simplicity that both densities are strictly positive.

a) Prove that the divergence is non-negative by using the inequality $\log(x) \leqslant x - 1$.

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