STATISTICAL MACHINE LEARNING (WS2022)
SEMINAR 8

Assignment 1. Let $x \in K^n$ be a random vector with components x_i from a finite set K. Consider two distributions $p(x)$ and $q(x)$ that factorise w.r.t. the components of x, i.e. $p(x) = \prod_{i=1}^n p_i(x_i)$ and similarly $q(x) = \prod_{i=1}^n q_i(x_i)$. Prove that their KL-divergence is the sum of KL-divergences of their marginal distributions, i.e.

$$D_{KL}(q(x) \parallel p(x)) = \sum_{i=1}^{n} D_{KL}(q_i(x_i) \parallel p_i(x_i)).$$

Assignment 2. Let X be a categorical random variable with values in $\mathcal{K} = \{1, \ldots, K\}$ and probabilities p_k, $k \in \mathcal{K}$. Solve the following optimisation tasks

$$\text{(a)} \quad \sum_{k \in \mathcal{K}} \alpha_k \log p_k \to \max_{\alpha}$$

$$\text{(b)} \quad \sum_{k \in \mathcal{K}} p_k \log \alpha_k \to \max_{\alpha}$$

$$\text{(c)} \quad \sum_{k \in \mathcal{K}} \alpha_k \log \frac{\alpha_k}{p_k} \to \min_{\alpha}$$

under the constraints $\alpha_k \geqslant 0, \forall k \in \mathcal{K}$ and $\sum_{k \in \mathcal{K}} \alpha_k = 1$.

Assignment 3. Let us consider a mixture of distributions from an exponential family, i.e.

$$p(x) = \sum_{k=1}^{K} \pi_k e^{\langle \phi(x), \eta_k \rangle - A(\eta_k)}$$

where $\eta = (\eta_1, \ldots, \eta_K)$ is the tuple of natural parameters and $\pi = (\pi_1, \ldots, \pi_K)$ is the tuple of mixture weights. Suppose you want to estimate η and π from a training set $\mathcal{T}^m = \{x^j \mid j = 1, \ldots, m\}$ by using the EM algorithm. In the E-step you will need to compute the optimal auxiliary variables

$$\alpha_x(k) = p\big(k \,|\, x; \eta^{(t)}, \pi^{(t)}\big), \quad \forall x \in \mathcal{T}^m$$

for the current estimate of η and π. In the M-step you will need to solve the optimisation task

$$\frac{1}{m} \sum_{x \in \mathcal{T}^m} \sum_{k=1}^{K} \alpha_x(k) \Big[\langle \phi(x), \eta_k \rangle - A(\eta_k) + \log \pi_k \Big] \to \max_{\eta, \pi}$$

a) Show that the task decomposes into independent optimisation tasks for η and π. Find the optimal tuple of mixture weights π.

b) Show that the optimisation task w.r.t. η further decomposes into independent tasks for each η_k. Show that each of them is an ML estimate for the respective η_k with the statistics

$$\psi_k = \frac{1}{m} \sum_{x \in \mathcal{T}^m} \alpha_x(k) \phi(x).$$

Assignment 4. Given a small training set $\mathcal{T}^m = \{x_i \in \mathbb{R} \mid i = 1, \ldots, m\}$ we want to estimate the mean of a normal distribution $\mathcal{N}(\mu, 1)$. We know that the unknown μ is close to μ_0. Therefore, we want to apply Bayesian inference and set the prior distribution for μ to be a normal distribution centred at μ_0, i.e. $p(\mu) = \mathcal{N}(\mu_0, 1)$.

Show that the posterior distribution $p(\mu \mid \mathcal{T}^m) \propto p(\mathcal{T}^m \mid \mu)\, p(\mu)$ is also a Gaussian. Find its center (expectation).