OPT13 - Information Theory
TP1: Entropy*

Stella Douka†, Guillaume Charpiat

March 8th, 2024

Problem 1 (Gibbs' inequality)**.** *Let* p *and* q *two probability measures over a finite alphabet* $\mathcal{X}$*. Prove that* $\mathrm{KL}(p \,\|\, q) \geqslant 0$

Hint: for a concave function f and a random variable X, we have the Jensen's inequality $\mathbb{E}\left[f(X)\right] \leqslant f\left(\mathbb{E}[X]\right)$. ln is a strictly concave function.

Problem 2 (Evidence Lower bound (ELBO))**.** *Prove the following inequality*[1]*:*

$$-\ln p(D) \leqslant -\mathbb{E}_{\theta\sim\beta}\left[\ln p(D|\theta)\right] + KL(\beta||\alpha) \tag{1}$$

where D *is a dataset,* $p(D)$ *is the probability of the dataset,* $p(D|\theta)$ *is the likelihood probability of the dataset given the model parameters* θ*,* β *is a distribution over the model parameters approximating the posterior distribution* $\pi(\theta) := p(\theta|D)$ *and* α *is the prior distribution over the model parameters.*

(a) Write down the natural logarithm of the Bayes' rule in an expanded form:

$$\pi(\theta) = \frac{p(D|\theta)\alpha(\theta)}{p(D)} \tag{2}$$

(b) Introduce a new density function β and rewrite the expression in terms of expectation w.r.t. β

(c) Use the Gibbs' inequality and write down the ELBO

(d) Interpret the ELBO in a machine learning framework

*https://www.lri.fr/~gcharpia/informationtheory/

†styliani.douka@inria.fr

[1] Further information can be found at http://www.yann-ollivier.org/rech/publs/mdltalks.php

Problem 3 (Entropy)**.** *Compute the differential entropy of the following distributions:*

(a) univariate Normal distribution

$$\mathcal{N}(x|\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right] \tag{3}$$

(b) multivariate Normal distribution

$$\mathcal{N}(x|\mu,C) = \frac{1}{\sqrt{(2\pi)^d|C|}} \exp\left[-\frac{1}{2}(x-\mu)^\top C^{-1}(x-\mu)\right] \tag{4}$$

where $x, \mu \in \mathbb{R}^d$ and C is a covariance matrix (assumed to be symmetric positive-definite).

Problem 4 (Mutual information)**.** *We are interested in computing the mutual information between a multivariate Normal distribution $\beta = \mathcal{N}(\mathbf{x}|\boldsymbol{\mu}, C)$ where $\mathbf{x}, \boldsymbol{\mu} \in \mathbb{R}^d$ and a product of identical univariate Normal distributions $\alpha = \prod_{i=1}^d \mathcal{N}(x_i|\mu,\sigma)$.*

(a) Express the KL divergence in terms of entropy and expectation w.r.t. β

(b) Compute the exact expression of $-\mathbb{E}_{x\sim\beta} \ln \alpha(x)$.

(c) Compute $KL(\beta||\alpha)$

(d) Suppose that $\mu_i = \mu$ and $C_{ii} = \sigma^2$ for all i. Simplify the previous expression.