ELBO loss function complexity term derivation for the correction of the EGUSPHERE-2023-1100 manuscript

Bent Harnist
bent.harnist@fmi.fi
Finnish Meteorological Institute, Helsinki, Finland

May 3, 2024

Preface

Dear editor and editing staff, this document shows the steps needed to arrive to the conclusion that the following statement in the latter part of the third (3) equation in manuscript EGUSPHERE-2023-1100 does not hold:

$$\underbrace{\overbrace{\mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})}[\log p(\boldsymbol{w})]}^{\text{prior}} - \overbrace{\mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})}[\log q(\boldsymbol{w} \mid \boldsymbol{\theta})]}^{\text{posterior}},}_{=-\mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})} D_{\mathrm{KL}}[q(\boldsymbol{w}|\boldsymbol{\theta})\|p(\boldsymbol{w})],\text{i.e., the complexity term}} \tag{1}$$

and should instead be:

$$\underbrace{\overbrace{\mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})}[\log p(\boldsymbol{w})]}^{\text{prior}} - \overbrace{\mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})}[\log q(\boldsymbol{w} \mid \boldsymbol{\theta})]}^{\text{posterior}},}_{=-D_{\mathrm{KL}}[q(\boldsymbol{w}|\boldsymbol{\theta})\|p(\boldsymbol{w})],\text{i.e., the complexity term}} \tag{2}$$

Derivation

Starting with:

$$\begin{aligned} &\mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})}[\log p(\boldsymbol{w})] - \mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})}[\log q(\boldsymbol{w} \mid \boldsymbol{\theta})] \\ &= -(\mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})}[\log q(\boldsymbol{w} \mid \boldsymbol{\theta})] - \mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})}[\log p(\boldsymbol{w})] \end{aligned} \tag{3}$$

We use the linearity of expected value and rules regarding the substraction of logarithms to transform the expression of Eq. 3 to:

$$\begin{aligned} &= -\mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})}[\log q(\boldsymbol{w} \mid \boldsymbol{\theta}) - \log p(\boldsymbol{w})] \\ &= -\mathbf{E}_{q(\boldsymbol{w}|\boldsymbol{\theta})}[\log \frac{q(\boldsymbol{w} \mid \boldsymbol{\theta})}{p(\boldsymbol{w})}] \end{aligned} \tag{4}$$

Using the definition of expected value, Eq. 4 is equal to

$$= -\int_{\boldsymbol{w} \in \boldsymbol{W}} q(\boldsymbol{w} \mid \boldsymbol{\theta}) \log \frac{q(\boldsymbol{w} \mid \boldsymbol{\theta})}{p(\boldsymbol{w})} d\boldsymbol{w} \tag{5}$$

where $\boldsymbol{W}$ is the probability space of sampled weights $\boldsymbol{w}$. Now, the Kullback–Leibler divergence between distributions A and B defined over the continuous probability space Ω is defined as $D_{\mathrm{KL}}[A \| B] = \int_{x \in \Omega} A(x) \log(\frac{A(x)}{B(x)}) \mathrm{d}x$. This definition, which when applied to Eq. 5 in reverse, yields:

$$= -D_{\mathrm{KL}}[q(\boldsymbol{w} \mid \boldsymbol{\theta}) \| p(\boldsymbol{w})] \tag{6}$$

which shows that Eq. 2 is the one that holds in the manuscript.

Best regards,
Bent Harnist (from the behalf of all authors)