SRVAE math

Etienne Bardet

May 2025

1 Introduction

2 Equation

Nous voulons maximiser $p(x)$ sachant les variables latentes u, z et y. La formulation de la log probabilité est la suivante :

$$p(x) = E_{q(w)}[\frac{p(x,w)}{q(w)}] \tag{1}$$

Commençons par réexprimer la loi jointe :

$$p(x,w) = p(x|y,u,z)p(z|y,u)p(y|u)p(u) \tag{2}$$

Nous pouvons négliger la dépendance en u pour x car les informations sont redondantes avec y. Nous avons également :

$$q(w|x) = q(z|y,x,u)q(u|x,y)q(y|x) \tag{3}$$

Or il n'y a aucun apport d'information de x sur u sachant x,y. Nous considérons également la transition $y|x$ comme déterministe. Ainsi, nous avons :

$$q(w|x) \approx q(z|y,x)q(u|y)q(y|x) \tag{4}$$

On a donc l'elbo suivant :

$$\begin{multline} log(p(x)) \geq E_{q(z|y,x)}(log(p(x|y,z))) + E_{q(z|y,x)q(u|y)}(log(p(z|y,u))) \\ + E_{q(u|y)}(log(p(y|u))) + E_{q(u|y)}(log(p(u))) - E_{q(z|y,x)q(u|y)}(log(q(z|y,x))) \\ - E_{q(u|y)}(log(q(u|y))) \end{multline} \tag{5}$$

On peut regrouper les termes suivants :

$$\begin{multline} E_{q(z|y,x)q(u|y)}(log(p(z|y,u))) - E_{q(z|y,x)q(u|y)}(log(q(z|y,x))) \\ = E_{q(u|y)}[q(z|y,x)log(\frac{p(z|y,u)}{q(z|y,x)})] \\ = -E_{q(u|y)}[q(z|y,x)log(\frac{q(z|y,x)}{p(z|y,u)})] \end{multline} \tag{6}$$

On retrouve donc une divergence de Kullback-Leibler :

$$-E_{q(u|y)}[q(z|y,x)log(\frac{q(z|y,x)}{p(z|y,u)}) = -E_{q(u|y)}(\mathcal{KL}(q(z|y,x)||p(z|y,u))) \quad (7)$$

Ici on peut simplifier la dépendance en y pour z car on considère que y n'apporte pas plus d'information que x, d'où :

$$-\mathcal{KL}(q(z|y,x)||p(z|y,u)) \approx -\mathcal{KL}(q(z|x)||p(z|y,u)) \quad (8)$$

De même on a :

$$- E_{q(u|y)q(y|x)}(log(q(u|y))) + E_{q(u|y)q(y|x)}(log(p(u))) = \\ - E_{q(y|x)}(q(u|y)log(\frac{q(u|y)}{p(u)}) = -E_{q(y|x)}(\mathcal{KL}(q(u|y)||p(u))) \quad (9)$$

Il reste donc :

$$ELBO = E_{q(z|y,x)}(log(p(x|y,z)))+E_{q(u|y)}(log(p(y|u)))-\mathcal{KL}(q(u|y)||p(u))-\mathcal{KL}(q(z|x)||p(z|y,u)) \quad (10)$$

Notons d'ailleurs qu'avec un décodeur Gaussien :

$$p(x|y,z) = \mathcal{N}(\hat{x},\gamma_1^2) \qquad p(y|u) = \mathcal{N}(\hat{y},\gamma_2^2) \quad (11)$$

Pour les décodeurs, nous avons donc en log :

$$log(p(x|y,z) = -log((2\pi)^{\frac{k}{2}}\gamma^k) + -\frac{||x-\hat{x}||_2^2}{2\gamma^2} \propto -k*log(\gamma) - \frac{||x-\hat{x}||_2^2}{2\gamma^2} \quad (12)$$

On a donc :

$$ELBO = -(k*log(\gamma_1)+E_{q(z|y,x)}(\frac{||x-\hat{x}||_2^2}{2\gamma_1^2}))-(k*log(\gamma_2)+E_{q(u|y)}(\frac{||y-\hat{y}||_2^2}{2\gamma_2^2})) \\ - E(\mathcal{KL}(q(u|y)||p(u))) - E(\mathcal{KL}(q(z|x)||p(z|y,u))) \quad (13)$$

Nous optimisons donc la loss suivante (négative Elbo) :

$$-ELBO = k*log(\gamma_1)+E_{q(z|y,x)}(\frac{||x-\hat{x}||_2^2}{2\gamma_1^2}))+k*log(\gamma_2)+E_{q(u|y)}(\frac{||y-\hat{y}||_2^2}{2\gamma_2^2})) \\ + E_{q(y|x)}(\mathcal{KL}(q(u|y)||p(u))) + E_{q(u|y)}(\mathcal{KL}(q(z|x)||p(z|y,u))) \quad (14)$$

Soit en posant :

$$MSE(x,\hat{x}) = \frac{1}{N}\sum_{n=1}^{N}(x_n - \hat{x}_n) \quad (15)$$

On obtient finalement :

$$- ELBO = k*log(\gamma_1) + \frac{k}{2\gamma_1}*MSE(x,\hat{x}) + k*log(\gamma_2) + \frac{k}{2\gamma_2}*MSE(y,\hat{y}) \\ + E_{q(y|x)}(\mathcal{KL}(q(u|y)||p(u))) + E_{q(u|y)}(\mathcal{KL}(q(z|x)||p(z|y,u))) \quad (16)$$

Avec k, la dimension des données.