SRVAE math

Etienne Bardet

May 2025

1 Introduction

Ce document a pour but de détailler les calculs formulant la loss d'un VAE dans un premier temps puis d'un VAE Conditionnel ("à deux étages") dans un second temps.

2 VAE

2.1 Évidence

Dans un VAE, nous cherchons à maximiser l'évidence, qui est la probabilité de retrouver notre données, conditionné sur les paramètres du modèle, soit :

$$p(x) = \int p(x|z)p(z)dz$$

2.2 Encodeur

Cependant, pour estimer $p(x|z)$, il nous faut connaitre $p(z|x)$ pour utiliser le théorème de Bayes.

Nous allons approcher cette distribution par un réseau de neurones qui fournira $q_\phi(z|x)$. Nous supposons ici que la distribution des données de l'espace latent est gaussienne, soit :

$$q_\phi(z|x) \hookrightarrow \mathcal{N}(\mu_\phi, \Sigma_\phi)$$

Avec Σ_ϕ une matrice de covariance diagonale (les l coefficients de sa diagonale étant l sorties de l'encodeur). μ_ϕ étant une autre sortie de l'encodeur, de même taille.

2.3 ELBO

Reformulons la formule de l'évidence :

$$p(x) = \int \frac{p(x,z)}{q_\phi(z|x)} q_\phi(z|x)dz$$

Le théorème du transfert nous permet d'obtenir l'égalité suivante

$$p(x) = \mathbb{E}_{q_\phi(z|x)}[\frac{p(x,z)}{q_\phi(z|x)}] \tag{1}$$

En passant, au log, nous avons

$$\log(p(x)) = \log(\mathbb{E}_{q_\phi(z|x)}[\frac{p(x,z)}{q_\phi(z|x)}])$$

Par concavité du logarithme, nous pouvons utiliser l'inégalité de Jensen, qui nous permet de borner la log-évidence

$$\log(p(x)) \geq \mathbb{E}_{q_\phi(z|x)}[\log(\frac{p(x,z)}{q_\phi(z|x)})] = \mathbb{E}_{q_\phi(z|x)}[\log(p(x,z)) - \log(q_\phi(z|x))] \tag{2}$$

2.3.1 $p(x, z)$

La formule d'une probabilité conditionnelle, nous donne pour $p(x, z)$

$$\log(p(x,z)) = \log(p(x|z)) + \log(p(z))$$

2.3.2 Attache aux données

Nous pouvons développer un terme d'attache aux données dans $\log(p(x|z))$ En effet, en partant sur le principe que notre décodeur est Gaussien, nous avons donc la loi suivante pour $p(x|z)$

$$p(x|z) \hookrightarrow \mathcal{N}(\hat{x}, \gamma^2\mathbb{I})$$

En utilisant la formule d'une distribution Gaussienne multivariée, nous pouvons donc développer le terme d'attache aux données comme suit

$$\log(p(x|z)) = -\frac{(x-\hat{x})^2}{2\gamma^2} - \log((2\pi)^{\frac{k}{2}}|\gamma^2\mathbb{I}|)$$

Nous ne pouvons pas jouer sur le terme d'échelle de la Gaussienne : $\frac{k}{2}\log(2\pi)$, nous allons donc utiliser la proportionnalité pour exprimer

$$\log(p(x|z)) \propto -\frac{(x-\hat{x})^2}{2\gamma^2} - k\log(\gamma) \tag{3}$$

À noter qu'ici, nous décidons de paramétrer la variance du décodeur par un paramètre γ qui est appris durant l'entrainement.

Note : k ici est une constante qui représente la dimension de x.

2.3.3 Reste de l'ELBO

Continuons par développer le reste de l'elbo dans notre équation. Il nous reste donc à développer les termes suivants :

$$\mathbb{E}_{q_\phi(z|x)}[\log(p(z)) - \log(q_\phi(z|x))]$$

Ces deux termes peuvent se regrouper pour donner notamment

$$-\mathbb{E}_{q_\phi(z|x)}[\log(\frac{q_\phi(z|x)}{p(z)})] = -\int q_\phi(z|x)\log(\frac{q_\phi(z|x)}{p(z)}) = -\mathcal{KL}(q_\phi(z|x)||p(z))$$

Notre ELBO peut donc s'écrire finalement de la façon suivante (on écrit le $-ELBO$ car nous allons l'optimiser en minimisant :

$$-ELBO = \frac{(x-\hat{x})^2}{2\gamma^2} + k\log(\gamma) + \mathcal{KL}(q_\phi(z|x)||p(z)) \tag{4}$$

3 VAE Conditionnel

De la même manière que dans la section précédente, nous voulons maximiser $p(x)$ Commencons par poser $w = [u, z, y]$. Nous avons alors

$$p(x) = \int p(x, w)dw = \int \frac{p(x, w)}{q(w|x)} q(w|x)dw \tag{5}$$

Le théorème du transfert donne alors

$$p(x) = \mathbb{E}_{q(w|x)}[\frac{p(x, w)}{q(w|x)}] \tag{6}$$

Commençons par réexprimer la loi jointe :

$$p(x, w) = p(x|y, u, z)p(z|y, u)p(y|u)p(u) \tag{7}$$

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

$$q(w|x) = q(z|y, x, u)q(u|x, y)q(y|x) \tag{8}$$

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 :

$$q(w|x) \approx q(z|y, x)q(u|y)q(y|x) \tag{9}$$

On a donc l'elbo suivant :

$$\begin{aligned}\log(p(x)) \geq\ & E_{q(z|y,x)}(\log(p(x|y,z))) + E_{q(z|y,x)q(u|y)}(\log(p(z|y,u))) \\ & + E_{q(u|y)}(\log(p(y|u))) + E_{q(u|y)}(\log(p(u))) - E_{q(z|y,x)q(u|y)}(\log(q(z|y,x))) \\ & - E_{q(u|y)}(\log(q(u|y)))\end{aligned} \tag{10}$$

On peut regrouper les termes suivants :

$$
\begin{aligned}
E_{q(z|y,x)q(u|y)}(\log(p(z|y,u))) - E_{q(z|y,x)q(u|y)}(\log(q(z|y,x))) \\
= E_{q(u|y)}[q(z|y,x)\log(\frac{p(z|y,u)}{q(z|y,x)})] \\
= -E_{q(u|y)}[q(z|y,x)\log(\frac{q(z|y,x)}{p(z|y,u)})] \quad (11)
\end{aligned}
$$

On retrouve donc une divergence de Kullback-Leibler :

$$
-E_{q(u|y)}[q(z|y,x)\log(\frac{q(z|y,x)}{p(z|y,u)}) = -E_{q(u|y)}(\mathcal{KL}(q(z|y,x)||p(z|y,u))) \quad (12)
$$

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 :

$$
-\mathcal{KL}(q(z|y,x)||p(z|y,u)) \approx -\mathcal{KL}(q(z|x)||p(z|y,u)) \quad (13)
$$

De même on a :

$$
\begin{aligned}
- E_{q(u|y)q(y|x)}(\log(q(u|y))) + E_{q(u|y)q(y|x)}(\log(p(u))) = \\
- E_{q(y|x)}(q(u|y)\log(\frac{q(u|y)}{p(u)}) = -E_{q(y|x)}(\mathcal{KL}(q(u|y)||p(u))) \quad (14)
\end{aligned}
$$

Il reste donc :

$$
ELBO = E_{q(z|y,x)}(\log(p(x|y,z))) + E_{q(u|y)}(\log(p(y|u))) - \mathcal{KL}(q(u|y)||p(u)) - \mathcal{KL}(q(z|x)||p(z|y,u)) \quad (15)
$$

Notons d'ailleurs qu'avec un décodeur Gaussien :

$$
p(x|y,z) = \mathcal{N}(\hat{x}, \gamma_1^2) \qquad p(y|u) = \mathcal{N}(\hat{y}, \gamma_2^2) \quad (16)
$$

Pour les décodeurs, nous avons donc en log :

$$
\log(p(x|y,z) = -\log((2\pi)^{\frac{k}{2}}\gamma^k) + -\frac{||x-\hat{x}||_2^2}{2\gamma^2} \propto -k*\log(\gamma) - \frac{||x-\hat{x}||_2^2}{2\gamma^2} \quad (17)
$$

On a donc :

$$
\begin{aligned}
ELBO = -(k*\log(\gamma_1) + E_{q(z|y,x)}(\frac{||x-\hat{x}||_2^2}{2\gamma_1^2})) - (k*\log(\gamma_2) + E_{q(u|y)}(\frac{||y-\hat{y}||_2^2}{2\gamma_2^2})) \\
- E(\mathcal{KL}(q(u|y)||p(u))) - E(\mathcal{KL}(q(z|x)||p(z|y,u))) \quad (18)
\end{aligned}
$$

Nous optimisons donc la loss suivante (négative Elbo) :

$$
\begin{aligned}
-ELBO = k*\log(\gamma_1) + E_{q(z|y,x)}(\frac{||x-\hat{x}||_2^2}{2\gamma_1^2})) + k*\log(\gamma_2) + E_{q(u|y)}(\frac{||y-\hat{y}||_2^2}{2\gamma_2^2})) \\
+ E_{q(y|x)}(\mathcal{KL}(q(u|y)||p(u))) + E_{q(u|y)}(\mathcal{KL}(q(z|x)||p(z|y,u))) \quad (19)
\end{aligned}
$$

Soit en posant :

$$MSE(x, \hat{x}) = \frac{1}{N} \sum_{n=1}^{N} (x_n - \hat{x}_n) \tag{20}$$

On obtient finalement :

$$\begin{aligned} -ELBO = k * \log(\gamma_1) + \frac{k}{2\gamma_1} * MSE(x, \hat{x}) + k * \log(\gamma_2) + \frac{k}{2\gamma_2} * MSE(y, \hat{y}) \\ + E_{q(y|x)}(\mathcal{KL}(q(u|y)||p(u))) + E_{q(u|y)}(\mathcal{KL}(q(z|x)||p(z|y, u))) \end{aligned} \tag{21}$$

Avec k, la dimension des données.

4 MoG

Admettons maintenant un mélange de gaussiennes :

$$p(u) = MoG(w, \mu, \sigma) = \sum_{i=1}^{Comp} w_i \mathcal{N}(\mu_i, \sigma_i^2) \tag{22}$$

Nous ne changeons qu'une chose : le prior de u. Qu'est ce que cela traduit pour l'ELBO ? Nous pouvons borner $\mathcal{KL}(p|| \sum_i w_i q_i) \leq \sum_i w_i \mathcal{KL}(p||q_i)$. Ainsi, nous avons donc une borne inf de l'ELBO qui est une borne inf de $p(x)$.