Bayes Formula

$$p(x|y) = \frac{p(x \cap y)}{p(y)} = \frac{p(y|x) \cdot p(x)}{p(y)}$$

Kullback-Leibler Divergence

The higher the probability of an event, the lower its information.

$I_p(x) \propto -p(x) \propto -\log p(x)$

Hence we set: $I_p(x) = -\log p(x)$

$$\Delta I = I_p(x) - I_q(x) = -\log p(x) + \log q(x)$$
$$= \log\left(\frac{q(x)}{p(x)}\right)$$

$$D_{KL}(p(x) \,\|\, q(x)) := \mathbb{E}_{\sim p(x)}[\Delta I]$$

$$= \int \Delta I \, p(x) \, dx = \int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx$$

<u>Properties</u>:

D_{KL} is not symmetric

$D_{KL}\left(p(x) \,\|\, q(x)\right) \geq 0$

Proof: Recall that $\log t \leq t-1$. We have:

$$-D_{KL}\left(p(x) \,\|\, q(x)\right) = -\int p(x) \log\left(\frac{p(x)}{q(x)}\right) dx$$

$$= \int p(x) \log\left(\frac{q(x)}{p(x)}\right) dx$$

$$\leq \int p(x)\left(\frac{q(x)}{p(x)} - 1\right) dx$$

$$= \int q(x)\, dx - \int p(x)\, dx = 1 - 1 = 0$$

Thus $D_{KL}\left(p(x) \,\|\, q(x)\right) \geq 0$

Log-likelihood

Soit $q(X, z)$ la distribution conjointe de X et images latentes z.

On a : $$q(x) = \int q(x, z)\, dz$$
$$= \int q(z \mid X = x)\, q(x)\, dz$$

Soit $p(z, X) = p(X, z)$ la distribution conjointe des images latentes z et data reconstruites X.

On a : $$p(x) = \int p(x, z)\, dz$$
$$= \int p(X = x \mid z)\, p(z)\, dz$$

Pour une bonne reconstruction, on veut $p(X) \simeq q(X)$.

→

On minimise la log-likelihood

$$\mathbb{E}_{q(X)}\left[\log p(X)\right] = \int q(x) \log p(x)\, dx$$

où $p(x) = \int p(X=x \mid z)\, p(z)\, dz$

Malheureusement, $p(x)$ est difficile à évaluer car requiert un parcours de tout l'espace latent z.

On recourt alors à des techniques bayesiennes...

Lemma:

$$\log p_\theta(x) \geq \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right] - \left.\begin{matrix} \\ D_{KL}\left(q_\phi(z|x) \,\|\, p_\theta(z)\right) \end{matrix}\right\} \mathcal{L}_\theta$$

and therefore, by taking $\mathbb{E}_{q_\phi(x)}$, we have:

$$\mathbb{E}_{q_\phi(x)}\left[\log p_\theta(x)\right] \geq$$

$$\mathbb{E}_{q_\phi(x)}\left[\mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right]\right] -$$

$$\mathbb{E}_{q_\phi(x)}\left[D_{KL}\left(q_\phi(z|x) \,\|\, p_\theta(z)\right)\right] := -\mathcal{L}(\phi, \theta).$$

Hence, minimizing $\mathcal{L}(\phi, \theta)$ maximizes the log-likelihood $\mathbb{E}_{q_\phi(x)}\left[\log p_\theta(x)\right]$.

The first and second term of $\mathcal{L}(\phi, \theta)$ are the <u>reconstruction</u> and <u>regularization</u> terms, respectively.

Proof:

$$D_{KL}\left(q_\phi(z|x) \,\|\, p_\theta(z|x)\right) =$$

$$-\int q_\phi(z|x) \log\left(\frac{p_\theta(z|x)}{q_\phi(z|x)}\right) dz = \text{(Bayes)}$$

$$-\int q_\phi(z|x) \log\left(\frac{p_\theta(x|z)\cdot p_\theta(z)}{q_\phi(z|x)\cdot p_\theta(x)}\right) dz =$$

$$-\left[-D_{KL}\left(q_\phi(z|x) \,\|\, p_\theta(z)\right) + \int q_\phi(z|x)\cdot \log\left(p_\theta(x|z)\right) dz - \int q_\phi(z|x)\cdot \log p_\theta(x)\, dz\right] =$$

$=1$

$$D_{KL}\left(q_\phi(z|x) \,\|\, p_\theta(z)\right) - E_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right] + \log p_\theta(x) \geq 0$$

since $D_{KL} \geq 0$

Therefore,

$$\log p_\theta(x) \geq E_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right] - D_{KL}\left(q_\phi(z|x) \,\|\, p_\theta(z)\right) \quad \blacksquare$$