TD-2 : CNN and VAE

Exercise 1 : Convolutional Neural Network

Question 1 :

Let consider a linear layer (with no bias). it takes as input $x \in \mathbb{R}^{10\times10\times3}$ an image and produce an intermediate representation $z^{(1)} \in \mathbb{R}^{10\times10\times20}$. How many trainable parameters (float) the model has (you can consider flattening input/ouput) ? What problem can we have if the input size of the image is high ?

Solution:

Question 2 :

Let consider a CNN 2 dimenssional layer, it takes as input $x \in \mathbb{R}^{10\times10\times3}$ an image and produce 20 features maps, the kernel or filter size are 3×3. How many trainable parameters (float) the model has ?

Solution:

Question 3 :

Considering the same architecture, what is the size of the features maps without considering padding ? If we consider a stride of 2 ?

Solution:

Exercise 2 : VAE

Question 1 :

Show that $\log(p(x)) \geq \mathbf{E}_{z\sim q_\psi(z|x)}\left[\log p_\theta(x|z)\right] - D_{KL}\left[q_\psi(z|x)\ ||p(z)\right]$

Solution: Start from $\log(p(x)) = log\int_z p(x,y)dz = \log\int_z p(x,y)\frac{q(z|x)}{q(z|x)}dz = \log\int_z q(z|x)\frac{p(x,y)}{q(z|x)}dz$
And notice that log is concave (so jensen) $log\int_z q(z|x)\frac{p(x,y)}{q(z|x)}dz \geq \int_z q(z|x)log\left(\frac{p(x,y)}{q(z|x)}\right)$ Then see the course

Question 2 :

What measures the KL divergence ? When it is equals to 0 ? Provide the associated formula for two distributions q and p

Solution: It is a disimilarity metrics over two distribution, it is minimum when the two distribution are equals.

$$KL(p(x)||q(x)) = \int_x p(x) \log\left(\frac{p(x)}{q(x)}\right)$$

Positivity

Notice that :

$$\int_x p(x) \log\left(\frac{p(x)}{q(x)}\right) = -\int_x p(x) \log\left(\frac{q(x)}{p(x)}\right)$$

By jensen's inequality we have (minus log is convex) :

$$\begin{aligned}\int_x p(x) \log\left(\frac{p(x)}{q(x)}\right) &\geq -\log\left(\int_x p(x)\frac{q(x)}{p(x)}\right)\\ &\geq -\log\left(\int_x q(x)\right)\\ &\geq -\log(1)\\ &\geq 0\end{aligned}$$

Question 3 :

Let consider $p(z)$ as the standard normal distribution (i.e $\mathcal{N}(0,1)$) and $q_\psi(z|x)$ a normal distribution parametrized by σ^2, μ, what is the expression of $D_{KL}\left(q_\psi(z|x)\ ||p(z)\right)$?

Solution:

- The PDF of q(z|x) is $\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}$
- The PDF of p(z) is $\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(z^2\right)}$

Thus,

$$
\begin{aligned}
KL\left(q(z|x)\ ||p(z)\right) &= \mathbb{E}_{z\sim q(z|x)}\left[\log\left(\frac{q(z|x)}{p(z)}\right)\right] \\
&= \mathbb{E}_{z\sim q(z|x)}\left[\log\left(\frac{\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}}{\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}(z^2)}}\right)\right] \\
&= \mathbb{E}_{z\sim q(z|x)}\left[\log\left(\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\right) - \log\left(\frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}\left(z^2\right)}\right)\right] \\
&= \mathbb{E}_{z\sim q(z|x)}\left[\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) + \log\left(e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\right) - \log\left(\frac{1}{\sqrt{2\pi}}\right) - log\left(e^{-\frac{1}{2}\left(z^2\right)}\right)\right] \\
&= \mathbb{E}_{z\sim q(z|x)}\left[\log\left(\frac{\sqrt{2\pi}}{\sigma\sqrt{2\pi}}\right) + \log\left(e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\right) - log\left(e^{-\frac{1}{2}\left(z^2\right)}\right)\right] \\
&= \mathbb{E}_{z\sim q(z|x)}\left[-\log\left(\sigma\right) + \log\left(e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\right) - log\left(e^{-\frac{1}{2}\left(z^2\right)}\right)\right] \\
&= \mathbb{E}_{z\sim q(z|x)}\left[-\log\left(\sigma\right)\right] + \mathbb{E}_{z\sim q(z|x)}\left[\log\left(e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\right)\right] - \mathbb{E}_{z\sim q(z|x)}\left[log\left(e^{-\frac{1}{2}\left(z^2\right)}\right)\right] \\
&= \int_z -\log\left(\sigma\right)q(z|x) + \mathbb{E}_{z\sim q(z|x)}\left[\log\left(e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\right)\right] - \mathbb{E}_{z\sim q(z|x)}\left[log\left(e^{-\frac{1}{2}\left(z^2\right)}\right)\right] \\
&= -\log\left(\sigma\right)\int_z q(z|x) + \mathbb{E}_{z\sim q(z|x)}\left[\log\left(e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\right)\right] - \mathbb{E}_{z\sim q(z|x)}\left[log\left(e^{-\frac{1}{2}\left(z^2\right)}\right)\right] \\
&= -\log\left(\sigma\right) + \mathbb{E}_{z\sim q(z|x)}\left[\log\left(e^{-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2}\right)\right] - \mathbb{E}_{z\sim q(z|x)}\left[log\left(e^{-\frac{1}{2}\left(z^2\right)}\right)\right] \\
&= -\log\left(\sigma\right) + \mathbb{E}_{z\sim q(z|x)}\left[-\frac{1}{2}\left(\frac{z-\mu}{\sigma}\right)^2\right] - \mathbb{E}_{z\sim q(z|x)}\left[-\frac{1}{2}\left(z^2\right)\right] \\
&= -\log\left(\sigma\right) - \frac{1}{2\sigma^2}\mathbb{E}_{z\sim q(z|x)}\left[(z-\mu)^2\right] + \frac{1}{2}\mathbb{E}_{z\sim q(z|x)}\left[\left(z^2\right)\right]
\end{aligned}
$$

Note that $\mathbb{E}_{z\sim q(z|x)}\left[(z-\mu)^2\right]$ is the definition of the variance
Note that :

$$
\begin{aligned}
\mathbb{E}_{z\sim q(z|x)}\left[z^2\right] &= \mathbb{E}_{z\sim q(z|x)}\left[(z-\mu)^2 - \mu^2 + 2\mu z\right] \\
&= \mathbb{E}_{z\sim q(z|x)}\left[(z-\mu)^2\right] - \mu^2 - \mathbb{E}_{z\sim q(z|x)}\left[2\mu z\right] \\
&= \sigma^2 - \mu^2 + 2\mu\mathbb{E}_{z\sim q(z|x)}\left[z\right] \\
&= \sigma^2 - \mu^2 + 2\mu^2 \\
&= \sigma^2 + \mu^2
\end{aligned}
$$

Thus,

$$\begin{aligned} KL\left(q(z|x)\ ||p(z)\right) &= -\log\left(\sigma\right) - \frac{1}{2\sigma^2}\sigma^2 + \frac{1}{2}\sigma^2 - \frac{1}{2}\mu^2 \\ &= -\log\left(\sigma\right) - \frac{1}{2} + \frac{1}{2}\sigma^2 + \frac{1}{2}\mu^2 \\ &= -\log\left(\sigma\right) - \frac{1}{2} + \frac{1}{2}\sigma^2 + \frac{1}{2}\mu^2 \\ &= -\frac{1}{2}\left(2\log\left(\sigma\right) + 1 - \sigma^2 - \mu^2\right) \\ &= -\frac{1}{2}\left(\log\left(\sigma^2\right) + 1 - \sigma^2 - \mu^2\right) \end{aligned}$$

Question 4 :

What is the loss associated to the ELBO (gradient descent) considering $p(z)$ as the standard normal distribution (i.e $\mathcal{N}(0,1)$) and $q(z|x)$ a normal distribution parametrized by σ^2, μ.

Solution: $\mathcal{L} = -\mathbb{E}_{z\sim q_\psi(z|x)}\left[p_\theta(x|z)\right] - \frac{1}{2}\left(\log\left(\sigma^2\right) + 1 - \sigma^2 - \mu^2\right)$

Question 5 :

How can we approximate $\mathbb{E}_{z\sim q_\psi(z|x)}\left[-log\left(p_\theta(x|z)\right)\right]$

Solution:
- Monte-Carlo to approximate the gradient (policy gradient)
- Use the reparametrization trick which also use monte-carlo to approximate z

Question 6 :

Let assume (in addition) that we consider reparametrization trick evaluating only one sample (not a good monte-carlo estimation for). We consider q_{enc} an MLP producing a representation of x, q_μ, q_σ layers producing the mean and the variance for distribution $q_\psi(z|x)$, p_{dec} the decoder function (modeling $p_\theta(x|z)$). Rewrite the loss function :

Solution: Let consider $\epsilon \sim \mathcal{N}(0,1)$ and that $\mu_x = q_\mu(q_{enc}(x))$ and $\sigma_x = q_\sigma(q_{enc}(x))$ then under above condition the loss can be written as :

$$\mathcal{L}(x) = -log\left(p_{dec}(\sigma_x\epsilon + \mu_x)\right) - \frac{1}{2}\left(\log\left(\sigma_x^2\right) + 1 - \sigma_x^2 - \mu_x^2\right)$$

Notice we can use any gradient descent algorithm to optimize this loss