

Modele generatywne 3: VAE

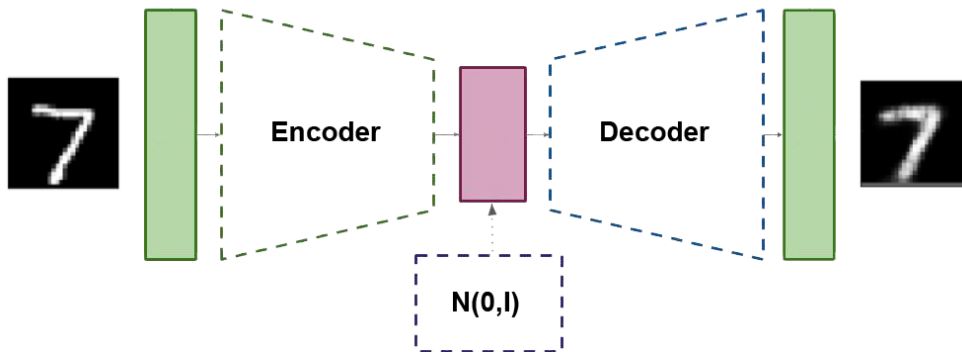
Jacek Tabor

13 października 2023

1 Model generatywny VAE

Ważniejszym, ale też bardziej skomplikowanym modelem niż WAE-MMD jest model VAE (Variational AutoEncoder), który jest historycznie pierwszy i ma także większy zakres stosowalności. Ogólna budowa VAE jest zbliżona do WAE-MMD i jest wspólna dla szerokiej klasy modeli generatywnych budowanych na bazie AutoEnkodera, to znaczy dane przechodzą przez przestrzeń latent, w której wymuszamy, by rozkład danych był gaussowski.

Podstawowym zadaniem modelu VAE jest maksymalizacja funkcji wiarygodności (log-likelihood). Ponieważ nie jest to możliwe bezpośrednio za pomocą jawnego wzoru w funkcji kosztu, używa się odpowiednio oszacowania górnego (tak zwane ELBO).



Rysunek 1: Ogólny wysokopoziomowy schemat VAE. Dane po przejściu do latent mają pochodzić z rozkładu normalnego, dodatkowo ma być dobra rekonstrukcja

Model VAE jest modelem probabilistycznym, to znaczy kod danego punktu x w przestrzeni nie jest jednoznacznie zdeterminowany, lecz jest rozkładem prawdopodobieństwa. Do jego opisu używa się rozkładów gaussowskich $N(\mu, \Sigma)$, co oznacza, że enkoder ma postać:

$$\mathbb{R}^N \ni x \xrightarrow{\mathcal{E}} N(\mu(x), \Sigma(x)).$$

Jeżeli chcemy teraz zakodować punkt x , to losujemy dowolny z z $\mathcal{E}(x)$, aby potem zdekodować bierzemy $\mathcal{D}(z)$
W praktyce najczęściej redukujemy się do $\Sigma(x)$ diagonalnego, czyli zadanego przez $\text{diag}(\sigma_1^2(x), \dots, \sigma_d^2(x))$.

Czyli w podstawowym VAE będziemy mieli: Enkoder wariacyjny $\mathcal{E} : \mathbb{R}^D \rightarrow \mathcal{N}(\mu(x), \text{diag}(\sigma^2(x)))$ zadana przez

- $\mathbb{R}^D \ni x \rightarrow \mu(x) \in \mathbb{R}^d$
- $\mathbb{R}^D \ni x \rightarrow \sigma(x) \in \mathbb{R}^d$ (czasami zadawane przez $\exp()$ by było nieujemne)

Dekoder

- $\mathcal{D} : \mathbb{R}^d \ni y \rightarrow \mathcal{D}y \in \mathbb{R}^D$

Innymi słowy enkoder jest zadawany za pomocą dwóch sieci, z których pierwsza zwraca $\mu(x)$ a druga $\sigma(x)$. Dekoder jest już deterministyczny i jest zbudowany analogicznie jak w modelu WAE. Podobnie jak w WAE-MMD, celem VAE jest wymuszenie, by rozkład danych przerzucony do latent space miał rozkład normalny.

Funkcja kosztu VAE: W modelu VAE chcemy zminimalizować wartość:

$$\|x - \mathcal{D}y\|^2 + D_{KL}(\mathcal{E}x, \mathcal{N}(0, I)) \text{ gdzie } y \sim \mathcal{E}x.$$

Pierwsza część odpowiada za błąd rekonstrukcji, zaś druga stara się zbliżyć do rozkładu normalnego. Formalnie jest to ograniczenie górne na $-\text{LogLikelihood}$ wygenerowania x w odpowiednim modelu probabilistycznym.

LICZENIE. Druga część jest łatwo liczalna – znamy wzór!

Natomiast jeżeli wyliczymy y , to nam zniknie gradient! W związku z tym, użyjemy REPARAMETRIZATION TRICK, aby przechodził gradient przez parametry:

- zamiast wylosować bezpośrednio z z $N(m, \text{diag}(\sigma_i^2))$,
- losujemy $\xi \sim N(0, I)$, i kładziemy

$$z = \mu + \sigma \odot \xi,$$

gdzie \odot to mnożenie componentwise,

- i widzimy teraz, że z jest zależne od parametrów.

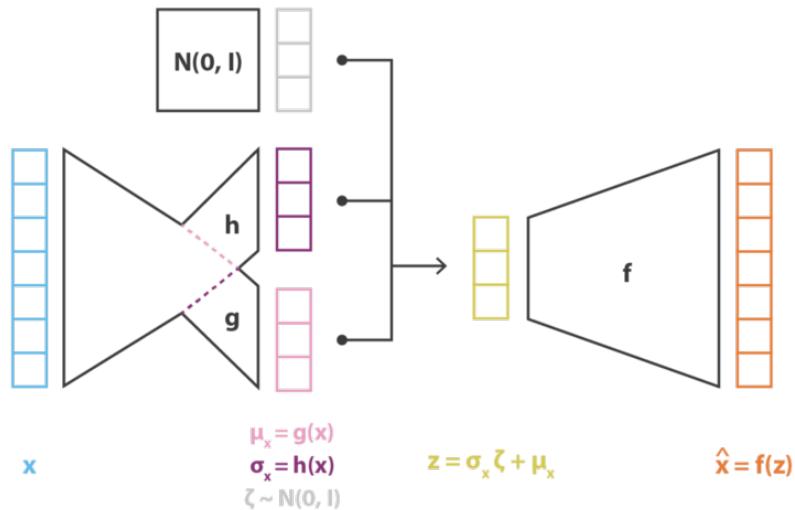
Finalnie funkcja kosztu VAE:

$$\|x - \mathcal{D}(\mu(x) + \sigma(x) \odot \xi)\|^2 + D_{KL}(\mathcal{E}x, N(0, I)) \text{ gdzie } \xi \sim N(0, I),$$

oraz

$$D_{KL}(\mathcal{E}x, N(0, I)) = \frac{1}{2} \left(\sum_i \sigma_i(x)^2 + \sum_i \mu_i(x)^2 - d - \sum_i \log(\sigma(x)_i^2) \right).$$

Różne VAE Są VAE dla związków chemicznych (ChemicalVAE), jest model VQVAE, gdzie mamy w latent dyskretne dane (używane w modelach dyfuzyjnych).



$$\text{loss} = C \|x - \hat{x}\|^2 + \text{KL}[N(\mu_x, \sigma_x), N(0, I)] = C \|x - f(z)\|^2 + \text{KL}[N(g(x), h(x)), N(0, I)]$$

Rysunek 2: Schemat i funkcja kosztu VAE.