Modele generatywne na chmurach punktów 3D

Jakub Zadrożny

Maj 2019

1 Podstawowe VAE

W pierwszej części projektu zaimplementowany został podstawowy autoenkoder wariacyjny. Dalej zakładamy, że dysponujemy zbiorem danych treninngowych

$$\mathcal{X} = \{x_i \in \mathbb{R}^d\}_{i \in I}$$

dla pewnego d – wymiaru danych.

Ponadto zakładamy, że dane są obserwacjami zmiennej losowej o rozkładzie następującej postaci

$$f(z, x; \theta) = f(z) f(x|z; \theta) \tag{1}$$

Dodatkowo niech

$$\begin{aligned} z &\sim \mathcal{N}(0, I_k) \\ x|z &\sim \mathcal{N}(\mu_x(z; \theta), \mu_\sigma(z; \theta) I_d) \end{aligned} \tag{2}$$

gdzie μ_x, μ_σ są skomplikowanymi obliczeniami wykonywanymi przez sieć neuronową sparametryzowaną przez θ.

1.1 ELBO

Naszym celem jest odtworzenie parametrów rozkładu generującego θ oraz rozkładu $f(z|x; \theta)$, który nazywamy *reprezentacją* danych generowanych przez proces opsiany w (1) oraz (2).

Niestety z powodu zastosowania skomplikowanych, nieliniowych transformacji dokładne odtworzenie rozkładu $f(z|x; \theta)$ jest niemożliwe. W tym celu wprowadzamy pewne przybliżenie tego rozkładu – nazwijmy je $g(z|x; \phi)$.

Niech $g(z|x;\theta)$ będzie gęstością rozkładu normalnego ze średnią $\rho_x(x;\phi)$ i wariancją $\rho_\sigma(x;\phi)$, gdzie ρ_x, ρ_σ są reprezentowane przez sieci neuronowe parametryzowane przez ϕ. Wtedy

$$\begin{aligned} D_{KL}(g(z|x;\phi)||f(z|x;\theta)) &= \mathbb{E}_{z\sim g(z|x;\phi)}\left[-\log\frac{f(z|x;\theta)}{g(z|x;\phi)}\right] = \\ &= \mathbb{E}_{z\sim g(z|x;\phi)}\left[-\log\frac{f(z|x;\theta)f(x;\theta)}{g(z|x;\phi)f(x;\theta)}\right] = \\ &= \mathbb{E}_{z\sim g(z|x;\phi)}\left[-\log\frac{f(z|x;\theta)f(x;\theta)}{g(z|x;\phi)}\right] + \mathbb{E}_{z\sim g(z|x;\phi)}\left[\log f(x;\theta)\right] = \\ &= \mathbb{E}_{z\sim g(z|x;\phi)}\left[-\log\frac{f(z,x;\theta)}{g(z|x;\phi)}\right] + \log f(x;\theta) \end{aligned} \tag{3}$$

Zatem

$$\log f(x;\theta) = D_{KL}(g(z|x;\phi)||f(z|x;\theta)) + \mathbb{E}_{z\sim g(z|x;\phi)}\left[\log\frac{f(z,x;\theta)}{g(z|x;\phi)}\right] \tag{4}$$

Ponieważ $D_{KL}(\cdot\,||\cdot\,) \geqslant 0$, więc

$$\begin{aligned} \log f(x;\theta) &\geqslant \mathbb{E}_{z\sim g(z|x;\phi)}\left[\log\frac{f(z,x;\theta)}{g(z|x;\phi)}\right] = \\ &= \mathbb{E}_{z\sim g(z|x;\phi)}\left[\log\frac{f(x|z;\theta)f(z)}{g(z|x;\phi)}\right] = \\ &= \mathbb{E}_{z\sim g(z|x;\phi)}\left[\log f(x|z;\theta)\right] - \mathbb{E}_{z\sim g(z|x;\phi)}\left[-\log\frac{f(z)}{g(z|x;\phi)}\right] = \\ &= \mathbb{E}_{z\sim g(z|x;\phi)}\left[\log f(x|z;\theta)\right] - D_{KL}(g(z|x;\phi)||f(z)) \end{aligned} \tag{5}$$

Zatem dla dowolnego rozkładu aproksymującego $g(z|x;\phi)$ otrzymujemy dolne ograniczenie na prawdopodobieństwo wygenerowania zaobserwowanych danych. Dlatego część wzoru po prawej stronie od ostatniej równości nazywamy ELBO (*evidence lower bound*). Ponadto pierwszy składnik odpowiada jakości rekonstrukcji obserwacji ze zmiennej ukrytej z, więc nazywany jest kosztem rekonstrukcji, natomiast drugi to odległość KL rozkładu aproksymującego $f(z|x;\theta)$ od naszego założenia na jego temat.

1.2 Zadanie optymalizacyjne

Chcemy znaleźć układ parametrów $< \theta, \phi >$, który daje najlepszą gwarancję na prawdopodobieństwo wygenerowania zaobserwowanych danych (ELBO). W tym celu posłużymy się lekko zmodyfikowanym algorytmem SGD. Naszym zadaniem jest znalezienie

$$\begin{aligned}\max_{\theta,\phi} \hat{\mathcal{L}}(\mathcal{X},\theta,\phi) &= \sum_{i\in I} \mathcal{L}(x_i,\theta,\phi) = \\ &= \sum_{i\in I} \Big(\mathbb{E}_{z\sim g(z|x_i;\phi)} \left[\log f(x_i|z;\theta) \right] - D_{KL}(g(z|x_i;\phi)||f(z)) \Big)\end{aligned} \tag{6}$$

Ponieważ bardziej naturalnym zadaniem jest minimalizowanie funkcji kosztu, to rozwiążemy równoważne zadanie znalezienia

$$\min_{\theta,\phi} -\hat{\mathcal{L}}(\mathcal{X},\theta,\phi) \tag{7}$$

Żeby posłużyć się algorytmem SGD musimy umieć wyliczać i różniczkować oba składniki funckji (L).

1.2.1 Koszt KL

Odległość KL dwóch rozkładów normalnych o następujących parametrach

$$\begin{aligned}\mathcal{N}_0 &\sim \mathcal{N}(\mu_0,\Sigma_0) \\ \mathcal{N}_1 &\sim \mathcal{N}(\mu_1,\Sigma_1)\end{aligned}$$

dla pewncyh $\mu_0,\ \mu_1 \in \mathbb{R}^k,\ \Sigma_0,\Sigma_1 \in \mathbb{R}^{k\times k}$, wynosi

$$D_{KL}(\mathcal{N}_0||\mathcal{N}_1) = \frac{1}{2}\left(\mathrm{tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1-\mu_0)^T\Sigma_1^{-1}(\mu_1-\mu_0) - k + \log\frac{\det\Sigma_1}{\det\Sigma_0} \right)$$

Ponieważ zakładamy, że $f(z)$ jest rozkładem $z \sim \mathcal{N}(0, I_k)$, więc

$$D_{KL}(g(z|x;\phi)||f(z)) = \frac{1}{2}\sum_{i=1}^{k} \Big(\rho_x(x;\phi)_i^2 + \rho_\sigma(x;\phi)_i^2 - \log(\rho_\sigma(x;\phi)_i^2) - 1 \Big) \tag{8}$$

Wzór (8) można wyliczać i różniczkować analitycznie.

1.2.2 Koszt rekonstrukcji

Drugiego składnika funckji $\mathcal{L}$, czyli kosztu rekonstrukcji, nie da się wyznaczyć analitycznie. Aby objeść ten problem, możemy metodą Monte Carlo oszacować wartość oczekiwaną przez średnią

$$\mathbb{E}_{z\sim g(z|x;\phi)}\left[\log f(x_i|z;\theta)\right] \sim \frac{1}{m}\sum_{i=1}^{m} -\log f(x|z_i;\theta)$$

gdzie $z_i \sim g(z|x;\phi)$. Taką wartość potrafimy już wyliczyć, ale nie potrafimy propagować gradientu do parametrów ϕ przez zaobserwowane wartości z_i.

Wprowadzimy repearametryzację zmiennych z_i – możemy zauważyć, że zmienna $z_i = \rho_x(x;\phi) + \epsilon_i\rho_\sigma(x;\phi)$ gdzie $\epsilon_i \sim \mathcal{N}(0, I_k)$ ma rozkład $g(z|x;\phi)$ a ponadto możemy propagować gradient do parametrów ϕ. Otrzymaliśmy zatem następujące przybliżenie na $\mathcal{L}$

$$\mathbb{E}_{z\sim g(z|x;\phi)}\left[\log f(x_i|z;\theta)\right] \sim \frac{1}{m}\sum_{i=1}^{m} -\log f(x|z_i = \rho_x(x;\phi) + \epsilon_i\rho_\sigma(x;\phi);\theta)$$

gdzie $\epsilon_i \sim \mathcal{N}(0, I_k)$

Ponieważ $x|z \sim \mathcal{N}(\mu_x(z;\theta), \mu_\sigma(z;\theta)I_d)$, więc

$$-\log f(x|z;\theta) = \sum_{i=1}^{d}\left(\frac{1}{2}\log(2\pi) + \log(\mu_\sigma(z;\theta)_i) + \frac{(x-\mu_x(z;\theta)_i)^2}{2\mu_\sigma(z;\theta)_i^2}\right)$$

jednak metryka ta niezbyt dobrze nadaje się do chmur punktów, ponieważ np. chcielibyśmy uznawać permutację punktów oryginalnej chmury za dobrą rekonstrukcję. Dlatego zamiast wyliczać *stricte* $\log f(x|z;\theta)$ skorzystamy ze zmodyfikowanego *Chamfer distance* danego wzorem

$$CD(\mathcal{X}_1, \mathcal{X}_2) = \sum_{x\in\mathcal{X}_1}\min_{y\in\mathcal{X}_2}||x-y||_2^2 + \sum_{x\in\mathcal{X}_2}\min_{y\in\mathcal{X}_1}||x-y||_2^2 \tag{9}$$

gdzie $\mathcal{X}_1$, $\mathcal{X}_2$ są zbiorami punktów wielowymiarowych. Ściślej mówiąc, możemy potraktować $\mu_x(z;\theta) \in \mathbb{R}^{3\cdot m}$ jako chmurę m punktów trójwymiarowych, oznaczmy ją $\hat{y}$. Ponadto dla $y \in \hat{y}$ niech $\sigma(y)$ oznacza 3-elementowy wektor wariancji $\mu_\sigma(z;\theta)$ utworzony ze składowych odpowiadających y. Za koszt rekonstrukcji przyjmiemy

$$\begin{aligned}\mathcal{L}_{rec}(\hat{x}|z,\theta,\phi) = &\sum_{x\in\hat{x}}\min_{y\in\hat{y}}\Big(-\log p_{y,\sigma(y)}(x)\Big) + \\ &+ \sum_{y\in\hat{y}}\min_{x\in\hat{x}}\Big(-\log p_{x,\sigma(y)}(y)\Big)\end{aligned} \tag{10}$$

gdzie $p_{v,s}(x)$ jest gęstością rozkładu normalnego o średniej v i macierzy kowariancji sI w punkcie x.

Po usunięcu stałych wyrazów można to zapisać jako

$$\begin{aligned}\mathcal{L}_{rec}(\hat{x}|z,\theta,\phi) = \sum_{x\in\hat{x}} \min_{y\in\hat{y}} \sum_{i=1}^{3} \left(\log(\sigma(y)_i) + \frac{(x_i - y_i)^2}{2\sigma(y)_i^2} \right) + \\ + \sum_{y\in\hat{y}} \min_{x\in\hat{x}} \sum_{i=1}^{3} \left(\log(\sigma(y)_i) + \frac{(x_i - y_i)^2}{2\sigma(y)_i^2} \right)\end{aligned} \tag{11}$$

W obecnej wersji modelu dla uproszczenia przyjęto, że $\mu_\sigma(z;\theta) = \alpha$ dla wszystkich z i niezależnie od parametrów θ (tzn. przyjęto stałą wariancję dla danych wyjściowych). Wtedy wzór (11) upraszcza się do

$$\begin{aligned}\mathcal{L}_{rec}(\hat{x}|z,\theta,\phi) &= \frac{1}{2\alpha^2} \left(\sum_{x\in\hat{x}} \min_{y\in\hat{y}} ||x-y||_2^2 + \sum_{y\in\hat{y}} \min_{x\in\hat{x}} ||x-y||_2^2 \right) = \\ &= \frac{1}{2\alpha^2} CD(\hat{x},\hat{y})\end{aligned} \tag{12}$$