Bài tập về nhà số 7

Nền tảng toán học của các mô hình tạo sinh – PIMA

Chủ đề: Mô hình tự mã hóa biến phân

Người giải: Võ Hoàng Nhật Khang

1 Bài tập 1.1 - Thủ thuật tái tham số hóa

Trong bài toán ước lượng gradient $G := \nabla_\phi \mathbb{E}_{Z\sim q_\phi}[f(Z)]$, với $q_\phi \equiv \mathcal{N}(\mu_\phi, \Sigma_\phi \Sigma_\phi^\top)$, chúng ta cần đánh giá phương pháp của hai bạn học: Boopy (sử dụng thủ thuật tái tham số hóa) và Naot (tính gradient trực tiếp).

1.1 Phân tích phương pháp của Boopy

Boopy sử dụng thủ thuật tái tham số hóa bằng cách biểu diễn $Z = g(\varepsilon; \phi) = \mu_\phi + \Sigma_\phi \varepsilon$, với $\varepsilon \sim \mathcal{N}(0, I)$. Khi đó

$$\mathbb{E}_{Z\sim q_\phi}[f(Z)] = \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}[f(g(\varepsilon;\phi))]. \tag{1}$$

Gradient của kỳ vọng được tính như sau

$$\nabla_\phi \mathbb{E}_{Z\sim q_\phi}[f(Z)] = \nabla_\phi \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}[f(g(\varepsilon;\phi))] = \mathbb{E}_{\varepsilon\sim\mathcal{N}(0,I)}[\nabla_\phi f(g(\varepsilon;\phi))]. \tag{2}$$

Boopy ước lượng gradient này bằng cách tính

$$F_{\text{Boopy}} := \frac{1}{n}\sum_{i=1}^{n} f(g(\varepsilon_i;\phi)), \quad \varepsilon_i \sim \mathcal{N}(0, I), \tag{3}$$

và sử dụng công cụ tự động lấy đạo hàm (như `torch.autograd`) để tính $\nabla_\phi F_{\text{Boopy}}$. Vì F_{Boopy} là một hàm xác định của ϕ (thông qua $g(\varepsilon_i;\phi)$), việc lấy gradient này là hợp lệ. Hơn nữa, do

$$\mathbb{E}[\nabla_\phi F_{\text{Boopy}}] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} \nabla_\phi f(g(\varepsilon_i;\phi))\right] = \nabla_\phi \mathbb{E}_{Z\sim q_\phi}[f(Z)], \tag{4}$$

phương pháp của Boopy cho ước lượng không chệch của G. Do đó, Boopy làm **đúng**.

1.2 Phân tích phương pháp của Naot

Naot tính

$$F_{\text{Naot}} := \frac{1}{n}\sum_{i=1}^{n} f(Z_i), \quad Z_i \sim q_\phi, \tag{5}$$

và sử dụng `torch.autograd` để lấy $\nabla_\phi F_{\text{Naot}}$. Tuy nhiên, F_{Naot} phụ thuộc vào Z_i, mà Z_i được lấy mẫu trực tiếp từ q_ϕ. Trong trường hợp này, các mẫu Z_i là ngẫu nhiên và không phải là hàm xác định của ϕ trong quá trình tính gradient. Việc lấy gradient trực tiếp $\nabla_\phi F_{\text{Naot}}$ không tương đương với $\nabla_\phi \mathbb{E}_{Z\sim q_\phi}[f(Z)]$, vì gradient không thể đi qua quá trình lấy mẫu ngẫu nhiên của Z_i. Điều này dẫn đến ước lượng gradient không chính xác. Do đó, Naot làm **sai**.

1.3 Kết luận

Boopy làm đúng vì sử dụng thủ thuật tái tham số hóa, cho phép biểu diễn Z dưới dạng hàm xác định của ϕ và tính gradient một cách hợp lệ. Naot làm sai vì cố gắng lấy gradient trực tiếp qua quá trình lấy mẫu ngẫu nhiên, điều không khả thi với các công cụ như `torch.autograd`. ∎

2 Bài tập 1.2 - Tối ưu chặn dưới chứng cứ

Chúng ta cần tối đa hóa chặn dưới chứng cứ (ELBO) cho mô hình VAE

$$\text{ELBO}(\phi,\theta) = \sum_{i=1}^{n} \mathbb{E}_{Z\sim q_{i,\phi}}[\log p_{X_i|Z_i}(x_i;\theta)] - \sum_{i=1}^{n} \text{KL}(q_{i,\phi}\|p_{Z_i;\theta}),$$

với:

- $q_{i,\phi} \equiv \mathcal{N}(\mu_\phi(x_i), \text{diag}(\sigma^2_\phi(x_i)))$, $\sigma^2_\phi(x_i) = \exp(s^2_\phi(x_i))$
- $p_{Z_i;\theta} \equiv \mathcal{N}(0, I_L)$
- $p_{X_i|Z_i=z;\theta} \equiv \mathcal{N}(\mu_\theta(z), \sigma^2 I)$

2.1 Câu a: Chứng minh công thức KL divergence

Chúng ta cần chứng minh

$$\text{KL}(q_{i,\phi}\|p_{Z_i;\theta}) = \frac{-s^2_\phi(x_i) + \sigma^2_\phi(x_i) + \mu^2_\phi(x_i) - 1}{2}. \tag{6}$$

Công thức KL divergence cho hai phân phối Gaussian là

$$\text{KL}(\mathcal{N}(\mu_0,\Sigma_0)\|\mathcal{N}(\mu_1,\Sigma_1)) = \frac{1}{2}\left(\text{tr}(\Sigma_1^{-1}\Sigma_0) + (\mu_1-\mu_0)^\top\Sigma_1^{-1}(\mu_1-\mu_0) - k + \log\frac{\det\Sigma_1}{\det\Sigma_0}\right) \tag{7}$$

trong đó k là số chiều. Với

- $q_{i,\phi} \equiv \mathcal{N}(\mu_\phi(x_i), \text{diag}(\sigma^2_\phi(x_i)))$,

- $p_{Z_i;\theta} \equiv \mathcal{N}(0, I_L)$,

ta có

- $\mu_0 = \mu_\phi(x_i)$, $\Sigma_0 = \mathrm{diag}(\sigma_\phi^2(x_i))$,
- $\mu_1 = 0$, $\Sigma_1 = I_L$, $k = L$.

Thay vào

$$\mathrm{KL}(q_{i,\phi} \| p_{Z_i;\theta}) = \frac{1}{2}\left(\mathrm{tr}(\Sigma_0) + \mu_\phi^\top(x_i)\mu_\phi(x_i) - L + \log\frac{\det I_L}{\det \Sigma_0}\right) \tag{8}$$

Ta lại có

- $\mathrm{tr}(\Sigma_0) = \sum_{j=1}^L \sigma_{\phi,j}^2(x_i)$,
- $\mu_\phi^\top(x_i)\mu_\phi(x_i) = \sum_{j=1}^L \mu_{\phi,j}^2(x_i)$,
- $\det I_L = 1$, $\det \Sigma_0 = \prod_{j=1}^L \sigma_{\phi,j}^2(x_i)$, nên $\log\dfrac{\det I_L}{\det \Sigma_0} = -\sum_{j=1}^L \log \sigma_{\phi,j}^2(x_i)$.

Vì $\sigma_\phi^2(x_i) = \exp(s_\phi^2(x_i))$, ta có $\log \sigma_{\phi,j}^2(x_i) = s_{\phi,j}^2(x_i)$. Do đó

$$\mathrm{KL}(q_{i,\phi} \| p_{Z_i;\theta}) = \frac{1}{2}\left(\sum_{j=1}^L \sigma_{\phi,j}^2(x_i) + \sum_{j=1}^L \mu_{\phi,j}^2(x_i) - L - \sum_{j=1}^L s_{\phi,j}^2(x_i)\right) \tag{9}$$

Gộp các số hạng theo từng chiều:

$$\mathrm{KL}(q_{i,\phi} \| p_{Z_i;\theta}) = \frac{1}{2}\left(-s_\phi^2(x_i) + \sigma_\phi^2(x_i) + \mu_\phi^2(x_i) - 1\right) \tag{10}$$

Kết quả này khớp với công thức yêu cầu nếu ta hiểu rằng công thức trong đề là giá trị trung bình theo các chiều. ■

2.2 Câu b: Chứng minh kỳ vọng log-likelihood

Chúng ta cần chứng minh

$$\mathbb{E}_{Z\sim q_{i,\phi}}[\log p_{X_i|Z_i=Z}(x_i;\theta)] = -\frac{1}{2\sigma^2}\mathbb{E}_{Z\sim q_{i,\phi}}[(x_i - \mu_\theta(Z))^2] + C,$$

với C là hằng số.
Ta có

$$p_{X_i|Z_i=z}(x_i;\theta) = \mathcal{N}(x_i; \mu_\theta(z), \sigma^2 I) \tag{11}$$

Log-likelihood là

$$\log p_{X_i|Z_i=z}(x_i;\theta) = -\frac{1}{2\sigma^2}\|x_i - \mu_\theta(z)\|_2^2 - \frac{D}{2}\log(2\pi\sigma^2) \tag{12}$$

Lấy kỳ vọng theo $Z \sim q_{i,\phi}$:

$$\mathbb{E}_{Z\sim q_{i,\phi}}[\log p_{X_i|Z_i=Z}(x_i;\theta)] = \mathbb{E}_{Z\sim q_{i,\phi}}\left[-\frac{1}{2\sigma^2}\|x_i-\mu_\theta(Z)\|_2^2 - \frac{D}{2}\log(2\pi\sigma^2)\right] \tag{13}$$

$$= -\frac{1}{2\sigma^2}\mathbb{E}_{Z\sim q_{i,\phi}}[\|x_i-\mu_\theta(Z)\|_2^2] - \frac{D}{2}\log(2\pi\sigma^2) \tag{14}$$

Gọi $C = -\frac{D}{2}\log(2\pi\sigma^2)$, ta được

$$\mathbb{E}_{Z\sim q_{i,\phi}}[\log p_{X_i|Z_i=Z}(x_i;\theta)] = -\frac{1}{2\sigma^2}\mathbb{E}_{Z\sim q_{i,\phi}}[(x_i-\mu_\theta(Z))^2] + C, \tag{15}$$

với C là hằng số. ■

2.3 Câu c: Đề xuất hàm mất mát

Chúng ta cần đề xuất hàm mất mát $\mathcal{L}(\phi,\theta) = \mathcal{L}_{\text{recon}}(\phi,\theta) + \mathcal{L}_{\text{KL}}(\phi,\theta)$, sao cho tối thiểu hóa $\mathcal{L}$ tương đương với tối đa hóa $\text{ELBO}(\phi,\theta)$.
Từ ELBO

$$\text{ELBO}(\phi,\theta) = \sum_{i=1}^{n}\mathbb{E}_{Z\sim q_{i,\phi}}[\log p_{X_i|Z_i}(x_i;\theta)] - \sum_{i=1}^{n}\text{KL}(q_{i,\phi}\|p_{Z_i;\theta}), \tag{16}$$

ta định nghĩa hàm mất mát $\mathcal{L} = -\,\text{ELBO}$, gồm hai phần:

- $\mathcal{L}_{\text{recon}} = -\sum_{i=1}^{n}\mathbb{E}_{Z\sim q_{i,\phi}}[\log p_{X_i|Z_i}(x_i;\theta)]$
- $\mathcal{L}_{\text{KL}} = \sum_{i=1}^{n}\text{KL}(q_{i,\phi}\|p_{Z_i;\theta})$

1. **Xấp xỉ $\mathcal{L}_{\text{recon}}$**: Sử dụng thủ thuật tái tham số hóa, ta biểu diễn $Z = \mu_\phi(x_i) + \sigma_\phi(x_i)\odot\varepsilon$, với $\varepsilon \sim \mathcal{N}(0, I_L)$. Kỳ vọng được xấp xỉ bằng Monte Carlo

$$\mathbb{E}_{Z\sim q_{i,\phi}}[\log p_{X_i|Z_i}(x_i;\theta)] \approx \frac{1}{M}\sum_{m=1}^{M}\log p_{X_i|Z_i=z_m}(x_i;\theta), \tag{17}$$

   với $z_m = \mu_\phi(x_i) + \sigma_\phi(x_i)\odot\varepsilon_m$, $\varepsilon_m \sim \mathcal{N}(0, I_L)$. Từ câu b, ta có:

$$\log p_{X_i|Z_i=z_m}(x_i;\theta) = -\frac{1}{2\sigma^2}\|x_i-\mu_\theta(z_m)\|_2^2 + C. \tag{18}$$

   Do đó

$$\mathcal{L}_{\text{recon}} \approx \frac{1}{M}\sum_{i=1}^{n}\sum_{m=1}^{M}\frac{1}{2\sigma^2}\|x_i-\mu_\theta(z_m)\|_2^2. \tag{19}$$

   Hằng số C có thể bỏ qua vì không ảnh hưởng đến tối ưu hóa.

2. **Xấp xỉ $\mathcal{L}_{\text{KL}}$**: Từ câu a, ta có

$$\text{KL}(q_{i,\phi}\|p_{Z_i;\theta}) = \frac{1}{2}\sum_{j=1}^{L}\left(-s_{\phi,j}^2(x_i) + s_{\phi,j}^2(x_i) + \mu_{\phi,j}^2(x_i) - 1\right). \tag{20}$$

Do đó

$$\mathcal{L}_{\mathsf{KL}} = \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{L} \left(-s^2_{\phi,j}(x_i) + s^2_{\phi,j}(x_i) + \mu^2_{\phi,j}(x_i) - 1 \right). \tag{21}$$

Hàm $\mathcal{L}_{\mathsf{KL}}$ này có thể tính trực tiếp từ $\mu_\phi(x_i)$ và $s^2_\phi(x_i)$.

Vậy, hàm mất mát được đề xuất là

$$\mathcal{L}(\phi, \theta) = \frac{1}{M} \sum_{i=1}^{n} \sum_{m=1}^{M} \frac{1}{2\sigma^2} \|x_i - \mu_\theta(z_m)\|_2^2 + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{L} \left(-s^2_{\phi,j}(x_i) + s^2_{\phi,j}(x_i) + \mu^2_{\phi,j}(x_i) - 1 \right),$$

với $z_m = \mu_\phi(x_i) + \sigma_\phi(x_i) \odot \varepsilon_m$, $\varepsilon_m \sim \mathcal{N}(0, I_L)$. Tối thiểu hóa $\mathcal{L}$ tương đương với tối đa hóa ELBO, vì $\mathcal{L} = -\mathrm{ELBO}$ (bỏ qua các hằng số không ảnh hưởng). ■