

Giới thiệu

Trong trường hợp các bài toán Machine Learning cổ điển, chúng ta phải tìm giá trị nhỏ nhất của hàm số f (cost / loss function) nào đó. Chiến lược chung là tìm các điểm cực tiểu toàn cục, hoặc trong một số bài toán khó cũng có thể chấp nhận nghiệm là cực tiểu địa phương thông qua đạo hàm. Đối với lớp các bài toán Quantum Machine Learning, ý tưởng chủ đạo vẫn được giữ nguyên.

Chiến thuật tìm các điểm cực tiểu được mô tả sau. Chúng ta có hàm f phụ thuộc biến $x \in X$ với x^* là cực tiểu toàn cục và $X = \{x^{(0)}, x^{(1)}, \dots, x^{(n)}\}$ là tập các thử nghiệm (ansatz) trong n vòng lặp, $x^{(0)}$ là giá trị khởi tạo, x^* là giá trị cần tìm, x có thể là vô hướng, vector, ma trận hoặc đại lượng nhiều chiều hơn. Ở đây ta cần đưa $x \approx x^*$ càng tốt và $x = x^*$ là tốt nhất.

Nhờ tính chất ngược hướng của đạo hàm (luôn hướng từ thấp lên cao, trong khi chúng ta đang cần x di chuyển từ cao xuống thấp), giá trị mới của x được cập nhật theo quy tắc sau:

$$x^{(t+1)} = x^{(t)} - \alpha \nabla f(x^{(t)})$$

với $\alpha \in \mathbb{R}$ là hệ số dùng để điều chỉnh tốc độ cập nhật (learning rate) và $\nabla f(x^{(t)}) = \left[\frac{\partial f(x^{(t)})}{\partial x_1} \quad \frac{\partial f(x^{(t)})}{\partial x_2} \quad \dots \right]$ là gradient của f tại $x^{(t)}$.

Tuy rõ ràng như vậy nhưng trong nhiều trường hợp f lại là hàm ẩn (black box) do không có thông tin chi tiết về f hoặc việc tính toán đạo hàm của f quá tốn kém.

1. Numerical Gradient

Để khắc phục nhược điểm này, chúng ta có một số phương pháp tính gián tiếp giá trị đạo hàm tại một điểm nào đó $f'(x)$ như sai phân hữu hạn, phương pháp này được chứng minh như sau:

Dùng khai triển Taylor với $f(x + \epsilon)$ và $f(x - \epsilon)$:

$$f(x + \epsilon) \approx f(x) + \frac{f'(x)(x+\epsilon-x)}{1!} + \dots = f(x) + f'(x)\epsilon + \dots \quad (1.1)$$

$$f(x - \epsilon) \approx f(x) + \frac{f'(x)(x-\epsilon-x)}{1!} + \dots = f(x) - f'(x)\epsilon + \dots \quad (1.2)$$

Lấy $\frac{(1.1)-(1.2)}{2\epsilon}$:

$$\frac{f(x+\epsilon)-f(x-\epsilon)}{2\epsilon} \approx f'(x) + \frac{f^{(3)}(x)}{6}\epsilon^2 + \dots = f'(x) + O(\epsilon^2) \quad (1.3)$$

Nếu $\epsilon \approx 0$, ta sẽ tính được gần đúng giá trị $f'(x)$ thông qua hai lần gọi hàm với độ lỗi $O(\epsilon^2)$. Đây còn có thể gọi là phương pháp bậc 0 vì không cần thông qua đạo hàm bậc cao hơn. Đối với tính

toán lượng tử thì $O(\epsilon^2)$ vẫn quá lớn để chấp nhận, ngoài ra định nghĩa đưa ra chỉ tính được xấp xỉ, có thể tiêu tốn nhiều thời gian tính toán trong trường hợp hàm xấu.

Tạm bỏ qua độ lỗi, ta có thể viết lại công thức (1.3) như sau:

$$f'(x) = r(f(x + \epsilon) - f(x - \epsilon)) \quad (1.4)$$

Ví dụ: khi áp dụng sai phân hữu hạn trên $f(x) = \sin(x)$ và $f'(x) = \cos(x)$, giả sử $f(x)$ là hàm ẵn, với $r = \frac{1}{2}$ và $\epsilon = \frac{\pi}{2}$, ta có:

$$\begin{aligned} f'(x) &= \frac{1}{2}(f(x + \frac{\pi}{2}) - f(x - \frac{\pi}{2})) \\ &= \frac{1}{2}(\sin(x + \frac{\pi}{2}) - \sin(x - \frac{\pi}{2})) \\ &= \frac{1}{2}(\cos(x) + \cos(x)) = \cos(x) \end{aligned}$$

2. Parameter shift

Bài toán đặt ra trong trường hợp này là làm như thế nào để tính được đạo hàm trên mạch lượng tử? Giả sử ta có chuỗi các cổng chứa tham số $U(\theta)$ phụ thuộc vào θ và quan sát \hat{B} , phụ thuộc vào quan sát có nghĩa là, chẳng hạn, nếu \hat{B} là ma trận Pauli

$$Z = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}.$$

Thì $\langle 0 | \hat{B} | 0 \rangle = 1$ và $\langle 1 | \hat{B} | 1 \rangle = -1$ tương ứng là trị riêng của pauli Z , $| 0 \rangle$ hay $| 1 \rangle$ trong ngữ cảnh này được gọi là trạng thái khởi tạo, trong nội dung tiếp theo mình sẽ kí hiệu chúng tổng quát là $| \psi \rangle$.

Hàm f cần tìm cực tiểu ở đây có dạng:

$$f(\theta) : R^m \rightarrow R^n := \langle 0 | U^\dagger(\theta) \hat{B} U(\theta) | 0 \rangle \quad (2.1) \quad (| \psi \rangle \text{ trong trường hợp này đơn giản là } | 0 \rangle)$$

Về mặt vật lý, chúng ta phải chạy mạch này nhiều lần để lấy trung bình k lần đo là xấp xỉ $f(0)$ với k là hyperparameter.

Để tiện cho việc đọc hiểu, chúng ta cần biết trước về một số chứng minh.

Chứng minh (2.1): Với B, C là hai toán tử bất kỳ, ta xét đại lượng $\langle \psi | B^\dagger \hat{Q} C | \psi \rangle + h.c$

$$= \langle \psi | B^\dagger \hat{Q} C | \psi \rangle + \langle \psi | C^\dagger \hat{Q} B | \psi \rangle$$

$$= \frac{1}{2} \langle \psi | (2B^\dagger C + 2C^\dagger B) \hat{Q} | \psi \rangle$$

$$\begin{aligned}
&= \frac{1}{2} \langle \psi | (B^\dagger B + B^\dagger C + C^\dagger B + C^\dagger C) \hat{Q} | \psi \rangle - \frac{1}{2} \langle \psi | (B^\dagger B - B^\dagger C - C^\dagger B + C^\dagger C) \hat{Q} | \psi \rangle \\
&= \frac{1}{2} \langle \psi | (B^\dagger + C^\dagger)(B + C) \hat{Q} | \psi \rangle - \frac{1}{2} \langle \psi | (B^\dagger - C^\dagger)(B - C) \hat{Q} | \psi \rangle
\end{aligned}$$

Chứng minh (2.2): Nếu m trị riêng của ma trận Hermit G có hai giá trị a, b phân biệt, thì ta có thể dời (shift) a, b về một cặp giá trị đối xứng $\pm r$ mà không làm mất tính tổng quan vì global phase không ảnh hưởng đến kết quả quan sát. Hệ quả là $G = r^2 \mathbb{I}$ với \mathbb{I} là ma trận đơn vị có kích thước tương ứng.

Ví dụ, xét ma trận Hermit

$$M = \begin{bmatrix} 0 & (1+i) \\ (1-i) & 0 \end{bmatrix}$$

có hai trị riêng $\pm \sqrt{2}$, ta có

$$M^2 = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} = 2\mathbb{I}$$

Chứng minh (2.3): Với \mathcal{G} là cổng unita và G là ma trận Hermit nào đó (có hai trị riêng phân biệt), ta có thể biểu diễn \mathcal{G} dưới dạng $\mathcal{G}(\mu) = e^{-i\mu G}$. Sử dụng khai triển Taylor:

$$e^{-i\mu G} = \sum_{k=0}^{\infty} \frac{(-i\mu)^k G^k}{k!}$$

Với $G = r^2 \mathbb{I}$ từ chứng minh (2.2):

$$\begin{aligned}
e^{-i\mu G} &= \sum_{k=0}^{\infty} \frac{(-i\mu)^{2k} (r^2 \mathbb{I})^{2k}}{2k!} + \sum_{k=0}^{\infty} \frac{(-i\mu)^{(2k+1)} (r^2 \mathbb{I})^{(2k+1)}}{(2k+1)!} \\
&= \mathbb{I} \sum_{k=0}^{\infty} \frac{(-1)^k (r\mu)^{2k}}{2k!} - \frac{i}{r} G \sum_{k=0}^{\infty} \frac{(-1)^k (r\mu)^{(2k+1)}}{(2k+1)!} \\
&= \mathbb{I} \cos(r\mu) - \frac{i}{r} G \sin(r\mu) \\
\rightarrow \mathcal{G}\left(\frac{\pi}{4r}\right) &= \frac{1}{\sqrt{2}} (\mathbb{I} - \frac{i}{r} G), \mathcal{G}\left(-\frac{\pi}{4r}\right) = \frac{1}{\sqrt{2}} (\mathbb{I} + \frac{i}{r} G)
\end{aligned}$$

3. Parameter shift trong trường hợp G có hai trị riêng phân biệt

Để đơn giản hóa, giả sử mạch đơn giản gồm tập con tham số $\mu \in \theta$ và duy nhất cổng có tham số \mathcal{G} phụ thuộc μ , đạo hàm trên toàn mạch là đạo hàm riêng của μ được ký hiệu là $\partial_\mu f(\mu)$. Ở đây mình dùng f thay cho $f(\mu)$ và \mathcal{G} thay cho $\mathcal{G}(\mu)$ vì cả hai đại lượng đều phụ thuộc vào μ . Lưu ý rằng trong trường hợp mạch phức tạp hơn, chúng ta có thể sử dụng quy tắc nhân trong đạo hàm để tính từng đạo hàm sau đó ghép chúng lại để cho ra kết quả cuối cùng.

Như vậy, ta có thể viết lại toàn bộ mạch dưới dạng $U(\theta) = V\mathcal{G}(\mu)W$, $\mathcal{G}(\mu)$ là cổng có tham số duy nhất, V, W là các bộ cổng phi tham số còn lại.

$$\begin{aligned} f(\mu) &= \langle 0 | U^\dagger \hat{B} U | 0 \rangle \\ &= \langle 0 | (V\mathcal{G}(\mu)W)^\dagger \hat{B} (V\mathcal{G}(\mu)W) | 0 \rangle \\ &= \langle 0 | W^\dagger \mathcal{G}^\dagger(\mu) (V^\dagger \hat{B} V) \mathcal{G}(\mu) W | 0 \rangle \quad (3.1) \end{aligned}$$

Hấp thụ V vào đại lượng quan sát \hat{B} và W vào trạng thái $|0\rangle$, ta thế $V^\dagger \hat{B} V = \hat{Q}$, $W | 0 \rangle = |\psi\rangle$ và $\langle 0 | W = \langle \psi |$ vào phương trình (3.1):

$$f = \langle \psi | \mathcal{G}^\dagger \hat{Q} \mathcal{G} | \psi \rangle$$

Khi lấy đạo hàm riêng của f đối với μ :

$$\begin{aligned} \partial_\mu f &= \partial_\mu \langle \psi | \mathcal{G}^\dagger \hat{Q} \mathcal{G} | \psi \rangle \\ &= \langle \psi | \mathcal{G}^\dagger \hat{Q} (\partial_\mu \mathcal{G}) | \psi \rangle + \langle \psi | (\partial_\mu \mathcal{G})^\dagger \hat{Q} \mathcal{G} | \psi \rangle \quad (3.2) \end{aligned}$$

Với $\mathcal{G}(\mu) = e^{-i\mu G}$ từ chứng minh (2.3), suy ra $\partial_\mu \mathcal{G}(\mu) = -iG e^{-i\mu G} = -iG\mathcal{G}(\mu)$ (3.3).

Thế (3.3) vào phương trình (3.2), ta được:

$$\partial_\mu f = \langle \psi | \mathcal{G}^\dagger \hat{Q} (-iG\mathcal{G}) | \psi \rangle + \langle \psi | (-iG\mathcal{G})^\dagger \hat{Q} \mathcal{G} | \psi \rangle$$

với $\langle \psi | \mathcal{G}^\dagger = \langle \psi' |$ và $\mathcal{G} | \psi \rangle = | \psi' \rangle$, $r \in \mathbb{R}$:

$$\begin{aligned} \partial_\mu f &= \langle \psi' | \hat{Q} (-iG) | \psi' \rangle + \langle \psi' | (-iG)^\dagger \hat{Q} | \psi' \rangle \\ &= r(\langle \psi' | \mathbb{I}^\dagger \hat{Q} (-\frac{i}{r}G) | \psi' \rangle + \langle \psi' | (-\frac{i}{r}G)^\dagger \hat{Q} \mathbb{I} | \psi' \rangle) \quad (3.4) \end{aligned}$$

Nếu n trị riêng của G chỉ có hai giá trị a và b với $a \neq b$, chúng ta có thể chuyển các trị riêng thành $\pm r$, vì pha toàn cục (global phase) không quan sát được.

Sử dụng chứng minh (2.1) trong phương trình (3.4) trong trường hợp $B = \mathbb{I}$ và $C = -\frac{i}{r}G$.

Lưu ý: G là ma trận Hermit ($G^\dagger = G$) nên $(-\frac{i}{r}G)^\dagger = \frac{i}{r}G$, ngoài ra thì $\mathbb{I}^\dagger = \mathbb{I}$.

$$\partial_\mu f = \frac{r}{2}(\langle \psi' | (\mathbb{I} - \frac{i}{r}G)^\dagger \hat{Q} (\mathbb{I} - \frac{i}{r}G) | \psi' \rangle + \langle \psi' | (\mathbb{I} + \frac{i}{r}G)^\dagger \hat{Q} (\mathbb{I} + \frac{i}{r}G) | \psi' \rangle) \quad (3.5)$$

Sử dụng kết quả từ chứng minh (2.3), $\mathcal{G}(\frac{\pi}{4r}) = \frac{1}{\sqrt{2}}(\mathbb{I} - \frac{i}{r}G)$ và $\mathcal{G}(-\frac{\pi}{4r}) = \frac{1}{\sqrt{2}}(\mathbb{I} + \frac{i}{r}G)$.

$$\partial_\mu f = \frac{r}{2}(\langle \psi' | \sqrt{2}\mathcal{G}^\dagger(\frac{\pi}{4r}) \hat{Q} \sqrt{2}\mathcal{G}(\frac{\pi}{4r}) | \psi' \rangle + \langle \psi' | \sqrt{2}\mathcal{G}^\dagger(-\frac{\pi}{4r}) \hat{Q} \sqrt{2}\mathcal{G}(-\frac{\pi}{4r}) | \psi' \rangle)$$

với $\langle \psi' | = \langle \psi | \mathcal{G}^\dagger$ và $|\psi'\rangle = \mathcal{G} |\psi\rangle$:

$$\partial_\mu f = r(\langle \psi | \mathcal{G}^\dagger \mathcal{G}^\dagger(\frac{\pi}{4r}) \hat{Q} \mathcal{G}(\frac{\pi}{4r}) \mathcal{G} | \psi \rangle + \langle \psi | \mathcal{G}^\dagger \mathcal{G}^\dagger(-\frac{\pi}{4r}) \hat{Q} \mathcal{G}(-\frac{\pi}{4r}) \mathcal{G} | \psi \rangle)$$

với $\mathcal{G}(\mu + s) = e^{-i(\mu+s)G} = e^{-i\mu G} e^{-isG} = \mathcal{G}(\mu) \mathcal{G}(s)$:

$$\begin{aligned} \partial_\mu f &= r(\langle \psi | \mathcal{G}^\dagger(\mu + \frac{\pi}{4r}) \hat{Q} \mathcal{G}(\mu + \frac{\pi}{4r}) | \psi \rangle + \langle \psi | (\mathcal{G}^\dagger(\mu - \frac{\pi}{4r}) \hat{Q} \mathcal{G}(\mu - \frac{\pi}{4r}) | \psi \rangle)) \\ &= r(f(\mu + \frac{\pi}{4r}) - f(\mu - \frac{\pi}{4r})) \quad (3.6) \end{aligned}$$

Như bạn có thể thấy, phương trình (3.6) trông khá giống phương trình (1.4) nhưng với $\epsilon = \frac{\pi}{4r}$ và định nghĩa chính xác, không xấp xỉ.

4. Bình luận

Phương pháp **parameter shift** có thể áp dụng trong nhiều trường hợp. Ví dụ như:

Trường hợp 4.1: nếu $G \in \frac{1}{2}\sigma$ với $\sigma = \{\sigma_x, \sigma_y, \sigma_z\}$ là 3 ma trận Pauli thì $r = \frac{1}{2}$ và $s = \frac{\pi}{2}$

Chứng minh: từ (3.6), vì $\frac{1}{2}\sigma_i$ có trị riêng là $\pm \frac{1}{2}$ nên $r = 1/2$ và $s = \frac{\pi}{4r} = \frac{\pi}{2}$ (4.1).

$$\begin{aligned} \text{Xét } \partial_\mu f &= \langle \psi | (\partial_\mu \mathcal{G}^\dagger) \hat{Q} \mathcal{G} | \psi \rangle + \langle \psi | \mathcal{G}^\dagger \hat{Q} (\partial_\mu \mathcal{G}) | \psi \rangle \\ &= \langle \psi | (\frac{i}{2}\sigma_i \mathcal{G}^\dagger) \hat{Q} \mathcal{G} | \psi \rangle + \langle \psi | \mathcal{G}^\dagger \hat{Q} (-\frac{i}{2}\sigma_i \mathcal{G}) | \psi \rangle \\ &= \frac{i}{2} \langle \psi | \mathcal{G}^\dagger (\sigma_i \hat{Q} - \hat{Q} \sigma_i) \mathcal{G} | \psi \rangle \\ &= \frac{i}{2} \langle \psi | \mathcal{G}^\dagger [\sigma_i, \hat{Q}] \mathcal{G} | \psi \rangle \quad (4.2) \end{aligned}$$

Với $[\sigma_i, \hat{Q}] = -i(\mathcal{G}^\dagger(\frac{\pi}{2}) \hat{Q} \mathcal{G}(\frac{\pi}{2}) - \mathcal{G}^\dagger(-\frac{\pi}{2}) \hat{Q} \mathcal{G}(-\frac{\pi}{2}))$ trong [Mitarai et al. \(2018\)](#), (4.2) trở thành:

$$\begin{aligned} &= \frac{1}{2}(\langle \psi | \mathcal{G}^\dagger (\mathcal{G}^\dagger(\frac{\pi}{2}) \hat{Q} \mathcal{G}(\frac{\pi}{2}) - \mathcal{G}^\dagger(-\frac{\pi}{2}) \hat{Q} \mathcal{G}(-\frac{\pi}{2})) \mathcal{G} | \psi \rangle) \\ &= \frac{1}{2}(\langle \psi | \mathcal{G}^\dagger (\mathcal{G}^\dagger(\frac{\pi}{2}) \hat{Q} \mathcal{G}(\frac{\pi}{2}) \mathcal{G} | \psi \rangle - \langle \psi | \mathcal{G}^\dagger (\mathcal{G}^\dagger(-\frac{\pi}{2}) \hat{Q} \mathcal{G}(-\frac{\pi}{2}) \mathcal{G} | \psi \rangle) \\ &= \frac{1}{2}(\langle \psi | \mathcal{G}^\dagger(\mu + \frac{\pi}{2}) \hat{Q} \mathcal{G}(\mu + \frac{\pi}{2}) | \psi \rangle - \langle \psi | \mathcal{G}^\dagger(\mu - \frac{\pi}{2}) \hat{Q} \mathcal{G}(\mu - \frac{\pi}{2}) | \psi \rangle) \\ &= \frac{1}{2}(f(\mu + \frac{\pi}{2}) - f(\mu - \frac{\pi}{2})) \quad (4.3) \end{aligned}$$

Từ (4.1) và (4.3), suy ra (3.6) trong trường hợp này đúng.

Trường hợp 2: nếu $G = r \vec{n} \sigma$ là tổ hợp tuyến tính của các toán tử Pauli σ với \vec{n} là vector 3 chiều bình thường, ta có thể thấy G có 2 trị riêng phân biệt và chứng minh (3.6) tương tự như trong trường hợp 1:

$$G = r \begin{bmatrix} n_X & n_Y & n_Z \end{bmatrix} \begin{bmatrix} \sigma_X \\ \sigma_Y \\ \sigma_Z \end{bmatrix} = r \begin{bmatrix} n_Z & (n_X - in_Y) \\ (n_X + in_Y) & n_Z \end{bmatrix}$$

có hai trị riêng phân biệt:

$$\lambda = \pm \sqrt{n_X^2 + n_Y^2 + n_Z^2}$$

Trường hợp 3: một số cổng cơ bản của Xmon qubit trong Google Cirq như $ExpW(\mu, \delta) = e^{-i\mu(\sigma_X \cos \delta + \sigma_Y \sin \delta)}$, $ExpZ(\mu) = e^{-i\mu\sigma_Z}$, $Exp11(\mu) = e^{-i\mu|11\rangle\langle 11|}$, ...

Tìm trị riêng của $ExpW(\mu, \delta)$:

$$G(\delta) = \sigma_X \cos \delta + \sigma_Y \sin \delta = \begin{bmatrix} 0 & (\cos \delta - i \sin \delta) \\ (\cos \delta + i \sin \delta) & 0 \end{bmatrix} = \begin{bmatrix} 0 & e^{-i\delta} \\ e^{i\delta} & 0 \end{bmatrix}$$

Giải phương trình $\det(G - \lambda \mathbb{I}) = 0$, suy ra $\lambda = \pm 1$, tương tự với $ExpZ(\mu)$ và $Exp11(\mu)$.

Tuy nhiên, các cổng dựa trên toán tử Pauli có thể không thỏa mãn điều kiện trên. Ví dụ như cổng biến đổi điều khiển bằng vi sóng trên kiến trúc siêu dẫn

$$\mathcal{G}(\mu) = e^{\mu(\sigma_X \otimes \mathbb{I} - b(\sigma_Z \otimes \sigma_X) + c(\mathbb{I} \otimes \sigma_X))}$$

có 4 trị riêng là nghiệm của phương trình:

$$\lambda^4 + 2\lambda^2(1 + b^2 + c^2) + (2b^2 - 2c^2 + (b^2 - c^2)^2 + 1) = 0.$$

Nhìn chung, **parameter shift** sẽ có một vài điểm vượt trội hơn phương pháp cũ, cụ thể là phương pháp sai phân hữu hạn như sau:

(1) Có thể áp dụng cho nhiều loại mạch khác nhau.

(2) Tính giá trị $\partial_\mu f(\theta)$ là chính xác với chỉ với hai lần đánh giá f trên mạch lượng tử.

(3) Độ dời (shift) ở đây được xác định, không nhất thiết là vô cùng bé như phương pháp sai phân hữu hạn. Do đó chúng ta có thể cực đại hóa khoảng cách giữa hai lần cập nhật trong không gian tham số.

Còn tiếp ...

Tài liệu tham khảo:

[1] https://pennylane.ai/qml/glossary/parameter_shift.html

[2] Schuld, Maria, et al. "Evaluating analytic gradients on quantum hardware." *Physical Review A* 99.3 (2019): 032331.

[3] Broughton, Michael, et al. "Tensorflow quantum: A software framework for quantum machine learning." *arXiv preprint arXiv:2003.02989* (2020).

[4] Mitarai, Kosuke, et al. "Quantum circuit learning." *Physical Review A* 98.3 (2018): 032309.