量子コンピュータの回帰問題への適応

name

2024 年 1 月 3 日

1　教師あり学習

機械学習の一種である教師あり学習について述べる。入力データ $\{\boldsymbol{x_i}\}$ と教師データ $\{f(\boldsymbol{x_i})\}$ を用意する。これらのデータを基にパラメータ $\boldsymbol{\theta}$ を調節して、$y_i \simeq f(\boldsymbol{x_i})$ となるような出力 $y_i = y(\boldsymbol{x_i}, \boldsymbol{\theta})$ を得るのが教師あり学習である。具体的には、教師データ $\{f(\boldsymbol{x_i})\}$ と出力 $y_i = y(\boldsymbol{x_i}, \theta)$ の近さを表す損失関数 $L = \sum_i \|f(\boldsymbol{x_i}) - y_i\|^2$ を最小化するパラメータ $\boldsymbol{\theta}$ を見つけることを指す。(ここら辺は、普通の機械学習をもうちょっと詳しくやってから加筆)

2　量子コンピュータによる教師あり学習

2.1　量子コンピュータの優位性

量子ビットの状態はベクトル、ゲートは行列によって記述され、1 量子ビットの自由度は 2 である。複数量子ビットの状態は、各量子ビットのテンソル積によって記述されるため、N 量子ビットを古典コンピュータで再現するとサイズ 2^N の行列を計算する必要がある。したがって、量子回路を古典コンピュータ上で再現し、観測の期待値を求めるには指数的な時間がかかる。対して、量子コンピュータを用いて観測の期待値を求める場合は、求める精度に応じたサンプリングを行えばよい。

つまり、指数オーダーの行列の計算が高速に行えることが期待されているのである。この特徴を、行列を多用するパーセプトロンの計算に用いるのが QCL(Quantum Curcuit Learning) である。

2.2　概要

量子コンピュータでの教師あり学習である QCL の概要を図 1 に示す。

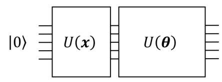

図 1: QCL の回路図

1. 入力データ $\{\boldsymbol{x_i}\}$ に対応する量子状態 $|\psi_{in}\rangle$ を生成する。ただし、$|0\rangle$ を $|\psi_{in}\rangle$ に変換するユニタリゲートを $U_{in}(\boldsymbol{x_i})$ とする。
2. パラメータ $\boldsymbol{\theta}$ を持つユニタリゲート $U(\boldsymbol{\theta})$(変分回路) を入力状態 $|\psi_{in}\rangle$ に作用させて出力状態 $|\psi_{out}\rangle$ をつくる。つまり、$|\psi_{out}\rangle = U(\boldsymbol{\theta})\,|\psi_{in}\rangle$
3. 適当な測定を行う。特に射影測定 $\{B_j\} \subset \{I, X, Y, Z\}^{\otimes N}$ を頻繁に用いる。
4. 測定量と学習モデルを関連付ける関数 F を用いて、出力 y_i を得る。つまり、$y(\boldsymbol{x_i}, \boldsymbol{\theta}) \equiv F(\{\langle B_j(\boldsymbol{x_i}, \boldsymbol{\theta})\rangle\})$
5. 損失関数 L が最小となるように $\boldsymbol{\theta}$ を決定する。

2.3　回転ゲート

変分回路 $U(\theta)$ のパラメータ θ は回転行列 $R_X(\theta), R_Y(\theta), R_Z(\theta)$ の回転角度として与えられる。パウリ行列 P で与えられる回転ゲート $R_P = \exp(-i\frac{\theta}{2}P)$ である。行列表示は、パウリ行列の以下

の性質を用いて、ネイピア数の行列乗の定義に従えば得られる。

$$P^n = \begin{cases} I & \text{(n is even)} \\ P & \text{(n is odd)} \end{cases} \tag{1}$$

$$R_P(\theta) = \cos(\frac{\theta}{2})I - i\sin(\frac{\theta}{2})P \tag{2}$$

2.4 入力状態の具体例 $|\psi_{in}\rangle$

N 量子ビットの量子回路における 1 入力の $x \in [-1, 1]$ の学習について述べる。入力状態の密度行列 $\rho_{in} = |\psi_{in}\rangle\langle\psi_{in}|$ は以下の式の状態とする。

$$\rho_{in}(x) = \frac{1}{2^N}\bigotimes_{i=1}^{N}\left[I + xX_i + \sqrt{1-x^2}Z_i\right] \tag{3}$$

具体的には、初期状態 $|0\rangle\langle 0|$ の各量子ビットに Y 軸回転ゲート $(\theta = \sin^{-1} x)$ を作用させる。

$$\begin{aligned} R_y(\theta) &= \exp\{-\frac{i}{2}\theta Y\} = \cos(\frac{\theta}{2})I - i\sin(\frac{\theta}{2})Y \\ \rho_{in} &= \bigotimes_{i=1}^{N} R_y(\theta)|0\rangle\langle 0| R_y(\theta)^\dagger \end{aligned} \tag{4}$$

密度演算子が ρ のとき、射影測定の測定量 P を測定すると、$\langle P\rangle = \mathrm{tr}(P\rho)$ である。密度行列 ρ_{in} の状態に対して測定量 $X^{\otimes N}$ を測定すれば、測定の期待値として x^N が得られる。観測の期待値は観測量、変分回路に依存しており、期待値は $x^2\sqrt{1-x^4}$ などの項の和で記述される。

2.5 観測の期待値とその勾配

2.5.1 観測の期待値

量子ビットが密度演算子 ρ で表せるとき、測定演算子が B である射影測定を行うと、その期待値は、$\langle B\rangle = \mathrm{tr}\,(B\rho)$ であることが導ける。

変分回路への入力状態の密度演算子を $\rho_{in} = |\psi_{in}\rangle\langle\psi_{in}|$、変分回路のパラメータの一つ θ_i に依存するゲートを $U_i(\theta_i)$ とする。また $U_m U_{m-1}\cdots U_{n+1}U_n \equiv U_{m:n}$ と定義する。

以上より、変分回路への入力状態の密度演算子を $\rho_{in} = |\psi_{in}\rangle\langle\psi_{in}|$ 変分回路の行列表示が $U_{l:1}$ のとき、変分回路の出力状態における、測定量 B の期待値は、以下の式で表せる。

$$\langle B\rangle = \mathrm{tr}\,(BU_{l:1}\rho_{in}U_{l:1}^\dagger) \tag{5}$$

2.5.2 観測の期待値の勾配

出力モデル $y_i(\boldsymbol{x}, \boldsymbol{\theta})$ をパラメータの一つ θ_i で微分した値は、損失関数の最小化の過程において重要である。そのため、観測の期待値 $\langle B\rangle = \mathrm{tr}(BU_{l:1}\rho_{in}U_{l:1}^\dagger)$ をあるパラメータ θ_i で微分した値を導出する。

まず、行列 $M = (a_{ij})$ について、行列の微分を以下の式 (6) で定義する。

$$\frac{dM}{d\theta_i} \equiv (\frac{da_{ij}}{d\theta_i}) \tag{6}$$

この微分について、以下の性質が行列の積の成分表示より求まる。

$$\frac{d}{d\theta_i}AB = \frac{dA}{d\theta_i}B + A\frac{dB}{d\theta_i} \tag{7}$$

$$\mathrm{tr}(\frac{dA}{d\theta_i}) = \frac{d}{d\theta_i}\,\mathrm{tr}(A) \tag{8}$$

$$\tag{9}$$

計算を簡単にするため、以下のように式を略記する、

$$\begin{cases} X = BU_{l:j+1} \\ Y = U_{j-1:1}\rho_{in}U_{j-1:1}^\dagger \\ Z = U_{l:j+1}^\dagger \\ [A, B] = AB - BA \end{cases} \tag{10}$$

以上の性質を用いて、観測値の期待を微分する。なお、$U_j = \exp(-\frac{i}{2}P_j\theta_j)$ とする。また、指数の行列乗の微分はテイラー展開を用いた定義より導出可

能である。

$$
\begin{aligned}
&\frac{\partial}{\partial\theta_i}\operatorname{tr}(BU_{l:1}\rho_{in}U_{l:1}^\dagger)\\
=&\operatorname{tr}\{X\frac{\partial}{\partial\theta_i}(U_jYU_j^\dagger)Z\}\\
=&\operatorname{tr}\{X(\frac{\partial U_j}{\partial\theta_j}YU_j^\dagger+U_j\frac{\partial Y}{\partial\theta_j}U_j^\dagger+U_jY\frac{\partial U_j^\dagger}{\partial\theta_j})Z\}\\
=&-\frac{i}{2}\operatorname{tr}\{XU_j[P_j,Y]U_j^\dagger Z\}\quad(\because U_j=\exp(-\frac{i}{2}P_j\theta_j))
\end{aligned}
\tag{11}
$$

ここで、以下の展開公式 (12) を用いる。

$$
[P_j,\rho]=i[U_j(\frac{\pi}{2})\rho U_j^\dagger(\frac{\pi}{2})-U_j(-\frac{\pi}{2})\rho U_j^\dagger(-\frac{\pi}{2})]
\tag{12}
$$

また、$U_j(\theta_0)U_j(\theta_1)=U_j(\theta_1)U_j(\theta_0)$ であることを用いる。回転ゲート $U_j(\theta_j)$ の固有ベクトルは、回転軸をあらわす状態ベクトルであるため、固有ベクトルが θ_j によらない。また、回転軸は 2 本の線形独立なベクトルで記述される (Z 軸. $|0\rangle, |1\rangle$)。したがって、同時対角化可能であるので、積について可換である。

これらの性質を用いれば、以下の式 (13) が求められる。式 (13) の意味としては、観測値 B をパラメータ θ_j で微分した値は、量子回路の特定の位置に回転ゲートを追加した回路の観測の期待値を求めればよいことが分かる。

$$
\begin{aligned}
\frac{\partial\langle B\rangle}{\partial\theta_i}=&\frac{1}{2}\operatorname{tr}(BU_{l:j+1}U_j(\frac{\pi}{2})U_{j:1}\rho_{in}U_{j:1}^\dagger U_j(\frac{\pi}{2})^\dagger U_{l:j+1}^\dagger)\\
&-\frac{1}{2}\operatorname{tr}(BU_{l:j+1}U_j(-\frac{\pi}{2})U_{j:1}\rho_{in}U_{j:1}^\dagger U_j(-\frac{\pi}{2})^\dagger U_{l:j+1}^\dagger)
\end{aligned}
\tag{13}
$$

3　数値計算

3.1　変分回路

変分回路の実装を図 2 に示す。図 2 では、観測が第 0 ビットの Z 基底観測である。

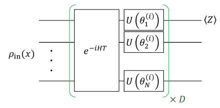

図 2: 変分回路 $U(\boldsymbol{\theta})$

まず、e^{-iH} について述べる。H は、次の式 (14) で表され、a_j, J_{jk} は $[-1,1]$ の一様乱数である。

$$
H=\sum_{j=1}^{N}a_jX_j+\sum_{j=1}^{N}\sum_{k=1}^{j-1}J_{jk}Z_jZ_k
\tag{14}
$$

e^{-iH} を古典コンピュータで計算する場合は固有値問題となり、効率的には行えない。また、e^{-iH} は量子ビットを量子もつれ状態にする。なお、T は 10 である。

$U(\theta_j^{(i)})(1\le j\le N, 1\le i\le D)$ は、ブロッホ球上の任意軸で任意の角度回転するゲートであり、次の分解 (15) を用いて実装される。

$$
U(\theta_j^{(i)})=R_j^X(\theta_{j1}^{(i)})R_j^Z(\theta_{j2}^{(i)})R_j^X(\theta_{j3}^{(i)})
\tag{15}
$$

3.2　回帰問題

図 3 のような、$\sin\pi x$ の入力と出力のデータを一定数与えて、$\sin\pi x$ を近似する量子回路のパラメータを求める。つまり、$\sin\pi x \approx \operatorname{tr}((Z\otimes I^{\otimes N-1})U_{l:1}(\boldsymbol{\theta})\rho_{in}(x)U_{l:1}^\dagger(\boldsymbol{\theta}))$ となるような、パラメータ $\boldsymbol{\theta}$ を求める。

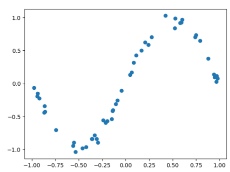

図 3: 教師データ $\sin(\pi x)$

入力状態を作成する回路 $U_{in}(x)$ は、式 (16) で表す。

$$
U_{in}(x)=\prod_j R_j^Z(\cos^{-1}x^2)R_j^Y(\sin^{-1}x)
\tag{16}
$$

今回用いる変分回路は、量子ビット数 $N=3$、図 2 の回路を $D=3$ 個つなげたものである。パラメータは基本軸回転ゲートの角度であり、回転ゲートが $3Nd$ 個あるので、パラメータは 27 個である。ランダムに初期化した際の、出力結果を図 4 に示す。

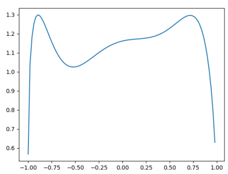

図 4: ランダムな初期値での出力

学習の手順としては、以下の手順で行う。

1. 27 個のパラメータ $\boldsymbol{\theta}$ を初期化する。
2. 教師データの入力 $\boldsymbol{x}_{train}$ より、現在のパラメータの出力 $\boldsymbol{y}$ を得る。
3. 現在のパラメータの出力と教師データの出力の差である損失関数の値を得る。
4. 勾配を用いる数理最適化手法の BFGS を用いて、2-3 を繰り返しながら、損失関数が最小になる $\boldsymbol{\theta}$ を得る。

なお、今回の数値計算に用いた python ライブラリである qulacs の仕様として、P 軸回転ゲートの角度の正負が反転している。ただし、教師データを含めた全てのパラメータの符号が反転しているため、学習には影響しない。

3.3 結果

学習の結果を図 5 に示す。損失関数の値が小さく、適切なフィッティングが行えていることが図から読み取れる。また、過学習が発生していないことも図から読み取れる。

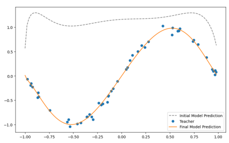

図 5: 最終的なモデルの出力

同じ回路に $\sin(2\pi x)$ を学習された結果を図 6 に示す。図から、適切な回帰が行えているとは言えない。

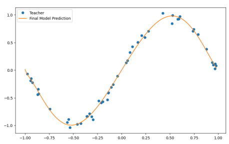

図 6: 近似できていない例

4 考察

4.1 1 量子ビットの場合の出力式

1 量子ビットの学習で得られる近似式を導出する。

1. $R_y(\sin^{-1}x), R_z(\cos^{-1}x^2)$ を x を含んだ行列表現として得る。
2. 最適化されたパラメータ $\boldsymbol{\theta_{opt}}$ を代入した各ゲートの行列表現を組み合わせて、変分回路 $U(\boldsymbol{\theta_{opt}})$ を行列表現で求める。
3. 観測値の期待値を求める。

4.1.1 入力状態 $|\psi_{in}\rangle$

変分回路に入力される状態 $|\psi_{in}\rangle = U_{in}(x)\,|0\rangle$ を求める。簡単化のため、次の略記を用いる。

$$\begin{cases} \alpha = \sqrt{1+x^2} \\ \beta = \sqrt{1-x^2} \\ \gamma = \sqrt{1+\beta} \\ \delta = \sqrt{1-\beta} \\ \epsilon = \dfrac{x}{|x|} \end{cases} \tag{17}$$

$\theta_1 = \sin^{-1} x, \theta_2 = \cos^{-1} x^2$ について、三角形の半角公式より次の関係が成り立つ。

$$\begin{cases} \sin \dfrac{\theta_1}{2} = \dfrac{\epsilon}{\sqrt{2}} \delta \\ \cos \dfrac{\theta_1}{2} = \dfrac{1}{\sqrt{2}} \gamma \\ \sin \dfrac{\theta_2}{2} = \dfrac{1}{\sqrt{2}} \beta \\ \cos \dfrac{\theta_2}{2} = \dfrac{1}{\sqrt{2}} \alpha \end{cases} \tag{18}$$

回転行列の性質である式 (2) より、変分回路への入力状態 $|\psi_{in}\rangle$ は、次の式 (19) で表せる。

$$|\psi_{in}\rangle = \frac{1}{2} \begin{pmatrix} \gamma(\beta + i\alpha) \\ -\epsilon\delta\overline{\beta + i\alpha} \end{pmatrix} \tag{19}$$

4.1.2 観測の期待値

量子ビット数が 1 のとき、変分回路の行列 M を式 (20) で記述する。各成分がパラメータ $\boldsymbol{\theta}$ に依存することに注意する。

$$M = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \tag{20}$$

今回は、Z 基底測定を行うため、測定の期待値 μ は式 (21) で表せる。

$$\mu = \langle\psi_{in}| M^\dagger Z M |\psi_{in}\rangle \tag{21}$$

式 (19)、(20)、(21) より、観測の期待値 μ は式 (22) で表せる。

$$\mu = A\sqrt{1 - x^2} + Bx^3 + C\sqrt{1 - x^4} \tag{22}$$

ただし、A, B, C は式 (23) で記述される実数であり、自由度は 3 であることが分かる。

$$\begin{cases} A = \dfrac{1}{2}(|a_{11}|^2 + |a_{22}|^2 - |a_{12}|^2 - |a_{21}|^2) \\ B = -\mathrm{Re}(a_{11}a_{12}^* + a_{21}a_{22}^*) \\ C = \mathrm{Im}(a_{11}a_{12}^* + a_{21}a_{22}^*) \end{cases} \tag{23}$$

A,B,C はパラメータ $\boldsymbol{\theta}$ に依存する。式 (22) より、学習とは目的関数に最も近い A,B,C を与えるパラメータを求めることであることが分かる。

4.2 N 量子ビットの場合の出力式

4.3 ニューラルネットワークとの比較

本実験で用いている変分回路は、以下の図 7 と等価である。

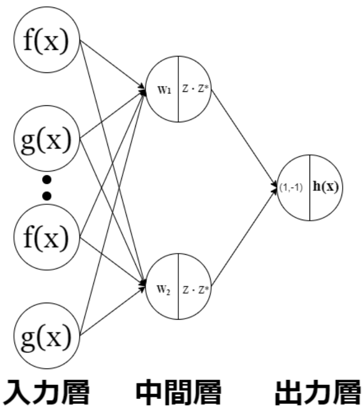

図 7: 変分回路と等価な NN

4.3.1 入力層について

入力層のノード数は、変分回路への入力される状態ベクトル $|\psi_{in}\rangle$ の次元に等しいため、量子ビット数が N のとき、ノード数は 2^N である。今回の実験では、$N = 3$ であるため、入力層のノード数は 8 である。$f(x), g(x)$ は、$U_{in}(x)$ に依存している各量子ビットの成分であり、本実験の場合は以下の式 (24) で表せる。古典 NN との大きな違いとして、$f(x), g(x)$ は複素関数であることが挙げられる。

$$\begin{cases} f(x) = \dfrac{1}{2}\sqrt{1 + \sqrt{1 - x^2}}(\sqrt{1 - x^2} + i\sqrt{1 + x^2}) \\ g(x) = \dfrac{x}{2|x|}\sqrt{1 - \sqrt{1 - x^2}}(\sqrt{1 - x^2} - i\sqrt{1 + x^2}) \end{cases} \tag{24}$$

4.3.2 中間層について

中間層のノード数は、測定が行われる量子ビットの個数の 2 倍であり、今回の実験では、第 0 ビット

のみに測定が行われるため、ノード数は 2 である。重み $\boldsymbol{w_i}(i = 1, 2)$ は、変分回路 $U(\theta)$ がユニタリ行列であることから、次の式 (25) を満たす。

$$\boldsymbol{w_i} \cdot \boldsymbol{w_j} = \delta_{ij} \tag{25}$$

また、通常の NN の非線形関数の対応として、ノードへの入力の絶対値の大きさの 2 乗をとっている。

4.3.3 出力層

出力層のノード数は、1 である。また、出力層の活性化関数は恒等写像である。重みは、観測量に依存している。本実験では観測量が Z 基底であるため、第 0 ビットの状態ベクトルが $\alpha |0\rangle + \beta |1\rangle$ で記述されるとき、期待値が $|\alpha|^2 - |\beta|^2$ である。したがって、重みは $(1, -1)$ である。

4.3.4 変分回路の自由度

今回の回路では、回路構成として、量子ビット数 $nqubit$ および $U(\boldsymbol{\theta})$ の内部で図 2 の回路を直列でつなげる個数である $cdepth$ を変更することが可能である。