

Read the fine print(Aaronson, 2015) 해설

김태원

최초 작성 : 2023년 9월 6일

최근 편집 : 2023년 9월 12일

양자기계학습^{quantum machine learning}은 HHL이라는 알고리즘에 바탕을 둔다. HHL은 Aram Harrow, Avinatan Hassidim, Seth Lloyd가 2008년에 공개한 알고리즘으로, 이후 나타난 양자기계학습 알고리즘은 대부분 HHL의 연장선에 있거나 HHL을 서브루틴으로 사용한다. HHL이 다루는 문제는 간단하다. 바로 **연립일차방정식**^{system of linear equations}을 푸는 것이다. $n \times n$ 실수 행렬 A 와 벡터 b 에 대해 아래 식에서 x 를 구하면 된다.

$$Ax = b.$$

HHL은 n 의 로그에 비례하는 연산만으로 $Ax = b$ 를 풀 수 있다고 주장한다. 고전적으로는 A 의 성분을 확인하는 작업에만 n^2 스텝이 요구되기 때문에 믿기지 않는 결과다. (어떻게 하는지는 나중에 살펴보고) 실제로 HHL은 $Ax = b$ 를 **로그시간**^{logarithmic time} 안에 풀지만, 몇 가지 조건이 있다.

첫째, $Ax = b$ 상의 벡터 $b = (b_0, \dots, b_{n-1})$ 가 양자컴퓨터의 메모리에 아무튼 재빠르게 로드되어야 한다. 아래 같은 양자상태를 미리 준비해야 하기 때문이다.

$$|b\rangle = \sum_{i=0}^{n-1} b_i |i\rangle = \begin{bmatrix} b_0 \\ b_1 \\ \vdots \\ b_{n-1} \end{bmatrix}$$

이를 이론상 가능케 하는 것이 이른바 “양자 RAM” 혹은 **QGRAM**이다. 즉 QGRAM은 수치들 b_0, \dots, b_{n-1} 를 양자중첩, 다시 말해 일차결합으로 한 번에 읽을 수 있도록 하는 장치다.

하지만 이런 QGRAM이 있다고 가정하더라도 별개 조건이 또 필요하다. 바로 벡터 b 가 **대체로 균일**^{relatively uniform}해야 한다는 것이다. 그러니까 어떤 b_i 가 나머지보다 지나치게 크면 안 된다.

제 1 장

서론

1.1 양자기계학습

저자Ewin Tang에 따르면 본고는 Kerenidis와 Prakash의 양자 추천 시스템 알고리즘quantum recommendation system algorithm이 고전 알고리즘에 비해 지수적인exponential 가속speedup¹을 보인다는 사실을 증명하려다 반증해 버린 결과물이다.

Kerenidis와 Prakash의 양자 추천 알고리즘은 대표적인 양자기계학습QML 알고리즘으로, QML의 진전은 2008년 Harrow, Hassidim, Lloyd가 제 이름을 따서 만든 연립방정식linear systems 풀이 알고리즘인 HHL 알고리즘에서 비롯한다. 다만 Aaronson이 *Read the Fine Print*에서 지적하듯 이들 QML 알고리즘이 고전 알고리즘에 대해 실질적으로 지수적인 가속을 보이냐는 문제는 명백하지 않다. 본고의 목적은 양자 추천 시스템 알고리즘과 같은 작업을 오직 다항적인 감속polynomially slower으로 수행할 수 있는 고전 알고리즘을 서술하여 이런 물음에 명백하게 아니라고 답할 수 있도록 하는 것이다.

1.1.1 양자 알고리즘이 작동하는 방법

양자 알고리즘은 입력 행렬 분해low-rank approximation of an input matrix로 표본을 취한다.

Low-rank approximation이 무엇이며, 애초에 **Rank**는 무엇인가?

- Rank-1 행렬은 영이 아닌 $m \times n$ 행렬로 m 벡터 u 와 n 벡터 v 의 외적 uv^T 으로 정의된다.

$$A = uv^T$$

¹지수적인 가속은 문제 해결 시간이 입력 크기에 대해 기하급수적으로 줄어드는 경우를 뜻한다.

- Rank-2 행렬은 두 Rank-1 행렬의 합 혹은 중첩이다.

$$A = uv^{\top} + wz^{\top}$$

그리하여 임의의 k 에 대해 Rank- k 행렬은 k 개의 Rank-1 행렬의 합 혹은 중첩이다. 그리고 low-rank matrix approximation은 주어진 행렬 A 를 Rank- k 행렬로 근사하는 과정이다.

입력 행렬에 대한 low-rank approximation으로 표본을 취한다는 말은 아래와 같다. 우선, 양자 위상 추정(quantum phase estimation) 알고리즘으로 단일 값을 추정하고 입력상의 단일 벡터를 배정한다.

Quantum phase estimation은 유니타리 연산자의 고윳값에 대한 근사다.

그런 다음 양자 사영(quantum projection) 과정은 이 정보를 통해 Kerenidis와 Prakash에 따르면 고전 알고리즘으로 이와 같은 분포로 표본을 생성하려면 입력 차원에 대해 선형인 시간이 요구된다. 하지만 저자가 보기에 입력 전체를 처리하지 않고 low-rank matrix approximation을 사용한다는 판단은 이상하다. 입력 데이터에 대해 강한 가정 내지 조건이 부여되는 것이기 때문이다.

제 2 장

정의

- 입력 데이터에 대한 기본적인 연산, 이를테면 덧셈, 곱셈, 읽기, 쓰기는 $O(1)$ 시간이 걸린다고 가정한다.
- $[n] := \{1, \dots, n\}$.
- $f \lesssim g$ 는 $f = O(g)$ 를 나타낸다.
- 행렬 A 에 대해, A_i 와 $A^{(i)}$ 는 각각 i 번째 행과 열을 나타낸다.
- $\|A\|_F$ 는 Frobenius norm을 나타낸다. Frobenius norm이란 $m \times n$ 행렬 A 의 노름으로, A 의 성분의 절댓값을 합하여 제곱근을 구한 것이다.

$$\|A\|_F = \sqrt{\sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2}$$

- $\|A\|_2$ 는 spectral norm을 나타낸다. 정방행렬 A 와 복소전치 A^\dagger 에 대해 spectral norm은 $A^\dagger A$ 의 최대 고윳값의 제곱근으로 정의된다.

$$\begin{aligned} \|A\|_2 &= (A^\dagger A \text{의 최대 고윳값})^{1/2} \\ &= \max_{|x|_2 \neq 0} \frac{|Ax|_2}{|x|_2} \end{aligned}$$

- 벡터 v 의 노름은 $\|v\|$ 으로 나타내고 항상 l^2 노름을 가리킨다.

$$\|x\| = \sqrt{\sum_{k=1}^n |x_k|^2} \quad \text{복소 벡터 } x \text{에 대해}$$

- 종종 $\|x - y\| \leq \epsilon$ 꼴의 행렬 부등식을 $\|E\| \leq \epsilon$ 에 대해 $x = y + E$ 로 나타낸다. E 는 오류에 대한 근사를 나타낸다.

- 행렬 $A \in \mathbb{R}^{m \times n}$ 에 대해 아래가 A 의 SVD가 되도록 한다.

$$A = U\Sigma V^T = \sum_{i=1}^{\min_{m,n}} \sigma_i u_i v_i^T$$

여기서 $U \in \mathbb{R}^{m \times n}$ 과 $V \in \mathbb{R}^{n \times n}$ 은 각각 열 $\{u_i\}_{i \in [m]}$ 과 $\{v_i\}_{i \in [m]}$ 을 지니는 유니타리 행렬로, 좌측과 우측 singular 벡터다. $\Sigma \in \mathbb{R}^{m \times n}$ 은 증가하거나 음이 아닌 $\sigma_i := \Sigma_{ii}$ 와 대각이다.

- 함수 ℓ 은 singular 벡터를 singular 벡터로 쪼개는 함수다.

$$\ell(\lambda) := \max\{i | \sigma_i \geq \lambda\}$$