Quantum Machine Learning Chapter 6

Ahn, Hyunjin

May 6, 2022

Research Institute of Mathematical Science, Seoul National University
yagamelaito@snu.ac.kr

Table of Contents

Table of Contents

Partial trace and reduced density operator

We consider the composite system which is the tensor product of the states of the component physical systems, for example, in pure bipartite states, A and B, the total Hilbert space $\mathcal{H}$ of the composite system is

$$|\Psi\rangle_{AB} = |\Psi\rangle_A \otimes |\Psi\rangle_B \in \mathcal{H} = \mathcal{H}_A \otimes \mathcal{H}_B.$$

Carrying out a measurement on a subsystem- an observable of system A with the operator M_A is a measurement of an observable of the composite system with the operator $M_A \otimes 1^B$. Analogously, measurements on the sub-system B is of the form $1^A \otimes M^B$.

Now, we let the pure state of the composite system

$$|\Psi\rangle = \sum_{a,b} \Psi_{ab} |e_a\rangle \otimes |f_b\rangle,$$

where the observable M^A of the subsystem A is measured.

Partial trace and reduced density operator

Then, we have

$$\begin{aligned}\langle M^A \otimes 1^B \rangle_\Psi = \langle \Psi | M^A \otimes 1^B \Psi \rangle &= \sum_{a_1, b_1} \sum_{a_2, b_2} \overline{\Psi_{a_2 b_2}} \Psi_{a_1 b_1} \langle e_{a_2} \otimes f_{b_2} | M^A e_{a_1} \otimes f_{b_1} \rangle \\ &= \sum_{a_1, b_1} \sum_{a_2, b_2} \overline{\Psi_{a_2 b_2}} \Psi_{a_1 b_1} \langle e_{a_2} | M^A e_{a_1} \rangle \langle f_{b_2} | f_{b_1} \rangle \\ &= \sum_{a_2, a_1, b} \overline{\Psi_{a_2 b}} \Psi_{a_1 b} \langle e_{a_2} | M^A e_{a_1} \rangle .\end{aligned}$$

On the other hand, we can also find same expectation value by using

$$\rho^A(\Psi) = \sum_{a_1, a_2, b} \overline{\Psi_{a_2 b}} \Psi_{a_1 b} | e_{a_1} \rangle \langle e_{a_2} | .$$

For this, we observe that $\rho^A(\Psi)$ is a density operator (i.e., self-adjoint, positive and has trace 1).

Partial trace and reduced density operator

Then, one has

$$\begin{aligned}\langle M^A\rangle_{\rho^A(\Psi)} &= \mathrm{tr}(\rho^A(\Psi)M^A) = \sum_a \langle e_a|\rho^A(\Psi)M^A e_a\rangle \\ &= \sum_{a,a_1,a_2,b} \overline{\Psi_{a_2 b}}\Psi_{a_1 b}\langle e_a|e_{a_1}\rangle\langle e_{a_2}|M^A e_a\rangle = \sum_{a_1,a_2,b} \overline{\Psi_{a_2 b}}\Psi_{a_1 b}\langle e_{a_2}|M^A e_{a_1}\rangle \\ &= \langle M^A \otimes 1^B\rangle_\Psi.\end{aligned}$$

Hence, $\rho^A(\Psi)$ describes a mixed state for the sub-system A.

Similarly, in the state $|\Psi\rangle$ of the composite system, the expectation of M^B for the sub-system B is as follows:

$$\langle 1^A \otimes M^B\rangle_\Psi = \sum_{b_1,b_2,a} \overline{\Psi_{ab_2}}\Psi_{ab_1}\langle f_{b_2}|M^B f_{b_1}\rangle = \langle M^B\rangle_{\rho^B(\Psi)},$$

where $\rho^B(\Psi) = \sum_{b_1,b_2,a} \overline{\Psi_{ab_2}}\Psi_{ab_1}|f_{b_1}\rangle\langle f_{b_2}|$.

Partial trace and reduced density operator

To generalize the concepts for $\rho^A(\Psi)$ and $\rho^B(\Psi)$, we need

Lemma

Assume that $\mathcal{H}^A$ and $\mathcal{H}^B$ are Hilbert spaces with respect to ONBs $\{|e_a\rangle\}$ and $\{|f_b\rangle\}$, respectively. Moreover, suppose $M \in L(\mathcal{H}^A \otimes \mathcal{H}^B)$ and let $M_{a_1 b_1, a_2 b_2}$ be the matrix of M in the ONB $\{|e_a \otimes f_b\rangle\}$ of $\mathcal{H}^A \otimes \mathcal{H}^B$ and let the operators $tr^B(M) \in L(\mathcal{H}^A)$ and $tr^A(M) \in L(\mathcal{H}^B)$ be given by

$$tr^B(M) = \sum_{a_1, a_2, b} M_{a_1 b, a_2 b} |e_{a_1}\rangle\langle e_{a_2}|, \quad tr^A(M) = \sum_{b_1, b_2, a} M_{b_1 a, b_2 a} |f_{b_1}\rangle\langle f_{b_2}|.$$

Then, $tr^B(M)$ and $tr^A(M)$ do not depend on the choice of the ONBs $\{|e_a\rangle\}$ and $\{|f_b\rangle\}$ and these are the unique operators such that

$$\begin{aligned} \forall M^A \in L(\mathcal{H}^A) &: \quad tr(M^A tr^B(M)) = tr((M^A \otimes 1^B)M), \\ \forall M^B \in L(\mathcal{H}^B) &: \quad tr(M^B tr^A(M)) = tr((1^A \otimes M^B)M). \end{aligned}$$

- We call $\mathrm{tr}^A(\mathrm{tr}^B)$ the partial trace over $\mathcal{H}_A(\mathcal{H}_B)$.

Partial trace and reduced density operator

This lemma defines a density operator ρ^A, which describes the state of the sub-system A when observed alone.

Theorem

Let $\rho \in D(\mathcal{H}^A \otimes \mathcal{H}^B)$ be the density operator describing the state of a composite system $\mathcal{H}^A \otimes \mathcal{H}^B$. Then

$$\rho^A(\rho) := tr^B(\rho)$$

is the uniquely determined density operator on $\mathcal{H}^A$, called by the reduced density operator, which describes the state in only the sub-system A is observed. For any observable M^A, it satisfies $\langle M^A \rangle_{\rho^A(\rho)} = \langle M^A \otimes 1^B \rangle_\rho$.

Let furthermore $\{|e_a\rangle\}$ be an ONB in $\mathcal{H}^A$ and $\{|f_b\rangle\}$ an ONB in $\mathcal{H}^B$ as well as $\rho_{a_1 b_1, a_2 b_2}$ be the matrix of ρ in the ONB $\{|e_a \otimes f_b\rangle\}$ in $\mathcal{H}^A \otimes \mathcal{H}^B$. Then, the matrix $\rho^A(\rho)$ in the ONB $\{|e_a\rangle\}$ is given by

$$\rho_A(\rho)_{a_1 a_2} = \sum_b \rho_{a_1 b, a_2 b}.$$

Table of Contents

Schmidt decomposition

To study the Schmidt decomposition, we again consider the pure state $|\Psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ and the corresponding density operator in the composite system as follows.

$$|\Psi\rangle = \sum_{a,b} \Psi_{ab} |e_a\rangle \otimes |f_b\rangle, \quad \rho^A(\Psi) = \sum_{a_1,a_2,b} \overline{\Psi_{a_2 b}} \Psi_{a_1 b} |e_{a_1}\rangle\langle e_{a_2}|.$$

Since $\rho^A(\Psi)$ is a self-adjoint and positive, there exists an ONB $\{|\tilde{e}_a\rangle\}$ in $\mathcal{H}^A$ consisting of eigenvectors of $\rho^A(\Psi)$ such that

$$\rho^A(\Psi) = \sum_a q_a |\tilde{e}_a\rangle\langle \tilde{e}_a|, \quad \text{where} \quad q_a \geq 0 \quad \text{are the eigenvalues.}$$

We note that there is a unitary U on $\mathcal{H}^A$ satisfying

$$|\tilde{e}_a\rangle = U|e_a\rangle = \sum_{a_1} |e_{a_1}\rangle \underbrace{\langle e_{a_1}|Ue_a\rangle}_{:=U_{a_1 a}}.$$

If we set

$$\tilde{\Psi}_{ab} := \sum_{a_1} U^*_{aa_1} \Psi_{a_1 b},$$

Schmidt decomposition

then it follows that

$$|\Psi\rangle = \sum_{a,b} \tilde{\Psi}_{ab} |\tilde{e}_a \otimes f_b\rangle,$$

which implies

$$\rho^A(\Psi) = \sum_{a_1,a_2,b} \tilde{\Psi}_{a_1 b} \overline{\tilde{\Psi}_{a_2 b}} |\tilde{e}_{a_1}\rangle\langle\tilde{e}_{a_2}|.$$

Thus,

$$\sum_b \tilde{\Psi}_{a_1 b} \overline{\tilde{\Psi}_{a_2 b}} = q_{a_2} \delta_{a_1 a_2}, \quad \text{moreover} \quad q_a = 0 \iff \tilde{\Psi}_{ab} = 0, \ \forall b.$$

For $q_a > 0$, we define the vectors

$$|\tilde{f}_a\rangle := \frac{1}{\sqrt{q_a}} \sum_b \tilde{\Psi}_{ab} |f_b\rangle \in \mathcal{H}^B.$$

$\implies$ The set of $|\tilde{f}_a\rangle$ is orthonormal.

Schmidt decomposition

Hence,

$$|\Psi\rangle = \sum_{a,b} \tilde{\Psi}_{ab} |\tilde{e}_a \otimes f_b\rangle = \sum_{q_a \neq 0} |\tilde{e}_a\rangle \otimes \sum_b \tilde{\Psi}_{ab} |f_b\rangle = \sum_{q_a \neq 0} \sqrt{q_a} |\tilde{e}_a \otimes \tilde{f}_a\rangle.$$

Here, we can drop the restriction $q_a \neq 0$, moreover, one can extend $|f_a\rangle$ to an ONB in $\mathcal{H}^B$.

Therefore, the Schmidt decomposition of $|\Psi\rangle \in \mathcal{H}^A \otimes \mathcal{H}^B$ is

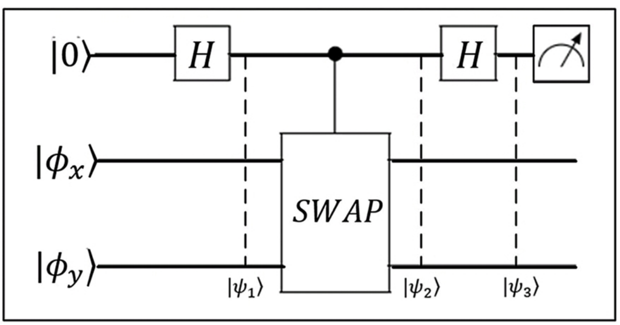
$$|\Psi\rangle = \sum_a \sqrt{q_a} |\tilde{e}_a \otimes \tilde{f}_a\rangle,$$

for the ONBs $|\tilde{e}_a\rangle$ and $|\tilde{f}_a\rangle$, where $\sqrt{q_a}$ is called by the Schmidt coefficient and the Schmidt rank is defined as the number of nonzero Schmidt coefficients. In particular, if the Schmidt rank is larger than 1, then the state is entangled , and if this is 1, then the state is separate.

Table of Contents

SWAP test

- SWAP test proposed by Buhman is very useful to measure the inner product between two quantum states.

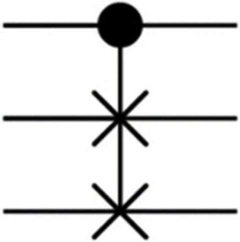


The quantum circuit for the swap test

- This circuit gives the information $|\langle\phi_x|\phi_y\rangle|^2$ when measuring the probability of $|0\rangle \rightarrow |1\rangle$.

SWAP test

- The circuit = an ancilla qubit + two qubit registers + two Hadamard gates + a controlled SWAP gate.



$$\equiv \textbf{CSWAP} \equiv \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

C-SWAP

Question. How to calculate the operator for C-SWAP?

SWAP test

If the ancilla qubit is superposed, and the $|\phi_x\rangle$ and $|\phi_y\rangle$ states are swapped, then by a direct calculation

1. $|\psi_1\rangle = H|0\rangle = \frac{1}{\sqrt{2}}(|0\rangle + |1\rangle)|\phi_x\rangle|\phi_y\rangle$.
2. $|\psi_2\rangle = \frac{1}{\sqrt{2}}\left(|0\rangle|\phi_x\rangle|\phi_y\rangle + |1\rangle|\phi_y\rangle|\phi_x\rangle\right)$.
3. $|\psi_3\rangle = \frac{1}{2}\left((|0\rangle + |1\rangle)|\phi_x\rangle|\phi_y\rangle + (|0\rangle - |1\rangle)|\phi_y\rangle|\phi_x\rangle\right)$.

SWAP test

Then, the probability being 1 of measuring the $|0\rangle$ is denoted by acceptance probability given by

$$\mathcal{P}_{\text{acceptance}} = \frac{1}{2}\left(1 - |\langle\phi_x|\phi_y\rangle|^2\right).$$

Question. How to verify this result?

Moreover, we note that

1. If $\langle\phi_x|\phi_y\rangle \to 0$, then $\mathcal{P}_{\text{acceptance}} \to \frac{1}{2}$.
2. If $|\phi_x\rangle = |\phi_y\rangle$, then $\mathcal{P}_{\text{acceptance}} = 0$.

Table of Contents

Kernel methods

- We consider a support vector machine (SVM) that finds a hyperplane (represented by normal vector w) that maximally separating vectors $\{v_1, \cdots, v_m\}$ by indicating labels $\{y_1, \cdots, y_m\}$, where $y_i = 1$ or -1.

- Training the SVM results in a linear prediction function

$$f(v_i) = \text{sign}(\langle w, v_i \rangle).$$

- Then, when

$$\langle w, v_i \rangle > 0 \quad \Longrightarrow \quad \{v_k \ : \ y_k = 1\} \quad \text{live on the half of one plane.}$$
$$\langle w, v_i \rangle < 0 \quad \Longrightarrow \quad \{v_k \ : \ y_k = -1\} \quad \text{live on the half of the other plane.}$$

Kernel methods

- Although, we did not define what space the vectors v, w live in, such space should have an associated inner product.

- In the Hilbert space, we take into account the following two algorithms for the SVM method.

1. Define the database $\{x_k\}$ to be a set of d-dimensional real vectors $\{x_1, \cdots, x_m\} = \mathcal{X} \subset \mathbb{R}^d$ so that we end up doing linear classification with hyperplane w and each label of x_i on $\mathcal{X}$.
2. Define the database $\{\phi_k\}$ to be a set of p-dimensional real vectors $\{\phi_1, \cdots, \phi_m\} = \Phi \subset \mathbb{C}^p$ so that we end up doing linear classification with hyperplane w' and each label of ϕ_i on Φ.

Kernel methods

- Then, we can consider a connection between the first algorithm and the second algorithm by using

$$\phi_i := \phi(x_i).$$

s.t. $\phi : \mathbb{R}^d \to \mathbb{C}^p$ is a feature map (Recall the feature map in Chapter 5).

- Unfortunately, ϕ is not always linear. Thus, the second algorithm is a nonlinear version of the first algorithm.

- Therefore,

$$f'(x_i) = \mathrm{sign}(\langle w', \phi(x_i) \rangle)$$

needs a decision boundary that is more explicit than the corresponding decision boundary $f(x_i) = \mathrm{sign}(\langle w, x_i \rangle)$.

Kernel methods

Question 1. What is quantum kernel method?
Answer. Constructing $\phi : x_i \mapsto \phi(x_i)$, where $\phi(x_i)$ live in quantum state space, typically, such as

$$\phi(x) = U(x)|0\rangle$$

for some unitary U that we can run on a quantum circuit.

Question 2. What is quantum kernel?
Answer. When processing our input data $\{x_i\}$ by evaluating inner products of the form $\langle \phi(x_i), \phi(x_j) \rangle = k(x_i, x_j)$, such k is called by kernel we define with respect to our data.

That is, the quantum kernel algorithm calculates a kernel matrix (Gram matrix).

THANK YOU FOR LISTENING