Quantum Mechanics as RNN

Ying Nian Wu
Department of Statistics
University of California, Los Angeles
ywu@stat.ucla.edu

This note explains the basics of quantum mechanics to people who are familiar with deep learning and neural networks. The key notion of deep learning or representation learning is embedding, such as word2vec and position embedding in transformer, as well as matrices operating on the vectors, such as query matrix, key matrix, value matrix, as well as embed matrix and unembed matrix for input layer and output layer. The key-value inner product in the softmax probability is also relevant. One may consider quantum mechanics a linear recurrent network (RNN) with a hidden layer, an embed layer, and an unembed layer.

1 Quantum mechanics as RNN model

The RNN model is as follows:

Schrödinger hidden layer:

$$h_{t+dt} = h_t + W h_t dt$$

Born unembed layer (output upon measurement):

$$p_t(o) \propto |\langle h_t, q_o \rangle|^2$$

Bohr embed layer (collapse upon measurement):

$$h_t = q_o$$

The Schrödinger layer tells us how the hidden vector h rotates. It is a residual form linear RNN or state space model, and the recurrent weight matrix

$$W = -iH$$

where H is the matrix version of Hamiltonian, to be explained later. $i = \sqrt{-1}$. The elements of vectors and matrices are complex numbers in general.

We require H to be symmetric (i.e., Hermitian, $H = H^*$, H^* is obtained by transposing H and taking complex conjugate of each element, $a + ib \to a - ib$), so that

$$h_{t+dt} = (I - iHdt)h_t$$

is a rotation, because

$$\begin{aligned}(I - iHdt)(I - iHdt)^* &= (I - iHdt)(I + iH^*dt) \\ &= I + O(dt^2)\end{aligned}$$

i.e., $(I - iHdt)$ is orthogonal.

We can also write the Schrödinger layer as

$$\frac{dh_t}{dt} = -iHh_t$$

The Born layer is the output layer, and $p_t(o)$ is the probability that the observer observes the state to be at value o upon measurement, and q_o is the query vector (or read-out vector, or embedding of o). You can interpret $\langle h_t, q_o\rangle$ as inner product between query and key vectors. It is a softmax layer except that we use square instead of exponential. Note that the elements of h_t and q_o are complex numbers, so $|\cdot|^2$ is complex square.

(q_o) are orthogonal vectors for the set of values (o), and they form a linear basis. If h_t is a unit vector under the rotation driven by $W = -iH$, then

$$|h_t|^2 = \sum_o |\langle h_t, q_o\rangle|^2 = 1$$

i.e.,

$$\sum_o p_t(o) = 1$$

This is why we use complex square in softmax, because the normalizing term $\sum_o |\langle h_t, q_o\rangle|^2$ is constant under rotation.

Upon output o, the hidden vector h_t collapses to q_o. This is the input layer, where classical observation o resets the hidden embedding h_t to q_o.

The hidden layer is the fundamental reality, and the embed and unembed layers provide interface with classical reality, which is a rendered display.

2 Hamiltonian of classical mechanics

The Newtonian mechanics can be rewritten in terms of Hamiltonian and Lagrangian. They are scalars, so they are much easier to work with than forces. Also they serve as stepping stones for generalization to quantum mechanics. The Lagrangian formulation is more convenient for calculation with special relativity, but the Hamiltonian formulation reveals the semantics of the mathematical language more clearly. So we shall only study Hamiltonian.

In classical mechanics, the Hamiltonian is the energy. For a free particle moving at velocity v,

$$H = \frac{1}{2}mv^2 = \frac{p^2}{2m}$$

where

$$p = mv$$

is the momentum, and $\frac{1}{2}mv^2$ is the kinetic energy. For simplicity, we can take $m = 1$ from now on.

If the particle is not free, we need to add potential energy $V(x)$, where x is position. For a harmonic oscillator (e.g., a particle attached to a spring), $V(x) = \frac{1}{2}x^2$, where again we omit terms in the coefficient of x^2 for simplicity. Then the Hamiltonian for the harmonic oscillator is

$$H = \frac{1}{2}(p^2 + x^2)$$

With Hamiltonian, you can write the differential equation for motion. We do not need to know the classical equation because quantum equation is more beautiful and in fact simpler.

3 Promote scalars to matrices

First we promote position x and momentum p to symmetric (i.e., Hermitian) matrices X and P, so that the Hamiltonian H as a function of X and P also becomes a matrix. For instance, for the harmonic oscillator,

$$H = \frac{1}{2}(X^2 + P^2)$$

Then we can use H to rotate h_t in the Schrödinger layer.

In the Born layer, we can choose to measure either X or P, or any other quantities, including H. These are called observables. Let us use O to denote an observable. O must be symmetric (i.e., Hermitian). Then we have the eigen decomposition

$$O = QDQ^*$$

where $D = \text{diag}(d_i)$ is diagonal matrix with (d_i) being the eigenvalues, and $Q = (q_i)$ is orthogonal matrix, with (q_i) being the eigenvectors. The vector q_i can be considered an embedding of the value d_i. Then in the Born layer, the probability the observer observes O at value d_i is

$$p_t(d_i) \propto |\langle h_t, q_i \rangle|^2$$

If we want to observe energy H, if $H = QDQ^*$, then the probability the observer observes H at d_i is according to the above probability.

Here the index i can be either discrete or continuous. For the harmonic oscillator, H has a discrete set of eigenvalues. This is the root of quantum phenomena.

4 About matrix or operator

For a square matrix M, we can use (i, i') to index its elements $(M_{i,i'})$, where i can be discrete, and the number of columns or rows can be infinite. More generally, the index can also be continuous, e.g., (x, x').

The matrix can be considered a verb, so that for a test vector v, which can be considered a noun, M transforms v to $u = Mv$.

In the case of continuous index, the element of v can be indexed by a continuous index such as x, then $v = (v(x))$ can be considered a test function of x, but we'd better not forget that v is also a vector whose x-th element is $v(x)$. Then M transforms v to $u = Mv$, and M is an operator that changes the function $v(x)$ to a new function $u(x) = (Mv)(x)$, where we can also think of Mv as a vector, whose x-th element is $u(x)$. In fact, we can discretize x to have equal spacing s, so that x is also a discrete index similar to i.

One example of operator is d/dx. For a test function $v(x)$, $d/dxv = u$, where u is the derivative of v, i.e., $u(x) = v'(x)$. If we discretize x with spacing s, then $u(x) = (v(x) - v(x - s))/s$, so d/dx is a matrix whose diagonal elements are $1/s$, and on each row x, the element in front of the diagonal element is $-1/s$. All the other elements are 0.

5 Quantization condition

In classical mechanics, the position x and the momentum p are two independent variables. In quantum mechanics, they become matrices (or operators) X and P respectively. For matrices, we know that XP is not equal to PX in general. The quantization condition is a constraint imposed on them:

$$XP - PX = iI$$

where I is the identity matrix (or operator).

This is a general condition assumed in quantum physics, replacing the old condition assumed by Bohr where the orbits of electron are quantized.

One special case is X is a diagonal matrix whose x-th diagonal element is x. Then for a test vector v, if $u = Xv$, then $u(x) = xv(x)$, i.e., X is an operator, so that Xv amounts to multiply x to $v(x)$. Then the eigen decomposition is

$$X = QDQ^* = IXI$$

where the x-column of Q (or I), let us call it q_x, is a one-hot vector whose x-th element is 1. This is the Dirac delta at x, i.e.,

$$q_x = \delta_x$$

$D = X$ is diagonal, whose x-th element is x. The one-hot vectors form a natural orthogonal basis I, and we call it the x-basis. In that basis, we can write momentum P as an operator

$$P = -i\frac{d}{dx}$$

As we discussed before, P can be viewed as a matrix if we discrete x with spacing s.

For a test function v,

$$\begin{aligned}[(XP - PX)v](x) &= -i(x\frac{d}{dx}v(x) - \frac{d}{dx}(xv(x))) \\ &= -i(xv'(x) - v(x) - xv'(x)) \\ &= iv(x)\end{aligned}$$

Thus $XP - PX = iI$, satisfying the quantization condition.

In the x-basis, the probability of observing X at x is

$$p_t(x) = |\langle h_t, \delta_x\rangle|^2$$

where $\langle h_t, \delta_x\rangle = h_t(x)$ is the wave function.

For the operator P, let its eigen decomposition be $P = QDQ^*$, where the eigenvectors are $(q_p = e^{ipx})$, i.e., the x-element of q_p is e^{ipx}.

$$Pq_p = -i\frac{d}{dx}e^{ipx} = pe^{ipx} = pq_p$$

so p is the eigenvalue of the eigenvector q_p. The vector q_p is the embedding of the observable value p of the momentum P. Thus the particle has a wave nature.

6 Uncertainty principle and transformation theory

Suppose in general we have eigen decompositions of X and P:

$$\begin{aligned}X &= QDQ^* \\ P &= Q'D'Q'^*\end{aligned}$$

If $Q = Q'$, then $XP = PX$. But we know $XP - PX = iI$. So Q and Q' are not the same. If the observer observes X with certainty according to the Born layer, then the normalized hidden vector in the Schrödinger layer, h, must be equal to an eigenvector q_i in Q, so that the observer observes value d_i with probability 1. However, in general, $h = q_i$ is not equal to any eigenvectors of Q'. Thus the observer cannot observe any values of P with certainty, and the probability the observer observes P at d'_j is $|\langle q_i, q'_j\rangle|^2 < 1$. This is the uncertainty principle.

For an operator M, let $u = Mv$, where u and v are vectors. We can think of u and v as nouns, and M as verb. If we change the basis to $Q = (q_i)$ with orthogonal vectors (q_i), then v becomes $v' = Q^*v$, or $v = Qv'$, i.e., v becomes v' viewed in Q. Similarly $u = Qu'$. Then $Qu' = MQv'$, and $u' = Q^*MQv' = M'v'$, i.e., viewed in Q, the verb M becomes Q^*MQ.

In Schrödinger picture, we fix the basis, and we let the state vector h_t rotate. In Heisenberg picture, we fix the state vector h_t, and we let the basis rotate. The two pictures are equivalent, as explained by Dirac's transformation theory.

7 Ladder operators

For the harmonic oscillator, the Hamiltonian is

$$H = \frac{1}{2}(X^2 + P^2)$$

with $XP - PX = iI$. We can factorize H as

$$\begin{aligned} H &= \frac{1}{2}(X + iP)(X - iP) - \frac{1}{2} \\ &= \frac{1}{2}(X - iP)(X + iP) + \frac{1}{2} \end{aligned}$$

Define

$$a = X + iP, \quad a^* = X - iP$$

then

$$H = \frac{1}{2}aa^* - \frac{1}{2} = \frac{1}{2}a^*a + \frac{1}{2}$$

Suppose q is an eigenvector of H with eigenvalue d, so that

$$Hq = dq$$

Then

$$\begin{aligned} H(a^*q) &= (\frac{1}{2}a^*a + \frac{1}{2})(a^*q) \\ &= \frac{1}{2}a^*aa^*q + \frac{1}{2}a^*q \\ &= \frac{1}{2}a^*(2H + 1)q + \frac{1}{2}a^*q \\ &= a^*dq + a^*q \\ &= (d + 1)(a^*q) \end{aligned}$$

That is,

$$q' = a^*q$$

is also an eigenvector of H with eigenvalue $d + 1$.

Similarly,

$$q'' = aq$$

is also an eigenvector of H with eigenvalue $d - 1$.

Thus we call (a^*) and (a) the ladder operators. (a^*) is the rising operator and (a) is the lowering operator. They were invented by Dirac.

Because $H = \frac{1}{2}(X^2 + P^2)$ is positive definite, the eigenvalues of H should be positive. Let q_0 be the eigenvector with the lowest eigenvalue. Then we must have

$$aq_0 = 0$$

Otherwise (aq_0) will be an eigenvector with an even smaller eigenvalue, which is a contradiction.

q_0 represents the ground state, whose corresponding eigenvalue is $\frac{1}{2}$ because

$$Hq_0 = (\frac{1}{2}a^*a + \frac{1}{2})q_0 = \frac{1}{2}q_0$$

Let

$$q_n = (a^*)^n q_0$$

then we get all the eigenvectors for $n = 0, 1, 2, \ldots$, and the eigenvalue of q_n is $(n + \frac{1}{2})$. There won't be eigenvectors with other eigenvalues, because repeatedly applying (a) to an eigenvector q will always end up at 0 vector, and there is only one ground eigenvector q_0 so that $aq_0 = 0$.

The harmonic oscillator underlies most of the quantum phenomena. (a^*) can also be considered the creation operator, and (a) the annihilation operator. q_0 is the state of nothing. $q_1 = a^*q_0$ is the state of one particle, and $q_2 = a^*q_1$ is the state of two particles, and so on.

8 Quantum field

For classical free field, it can be written as

$$\phi(x) = \int [a^*(k)e^{-ikx} + a(k)e^{ikx}]dk$$

where $a^*(k)$ and $a(k)$ are conjugate complex scalars.

For quantum field, $a^*(k)$ and $a(k)$ become operators. $a^*(k)$ is the creation operator that creates a particle with momentum k, and $a(k)$ is the annihilation operator that annihilates a particle with momentum k. The ground state q_0 represents the state of "nothingness". Thus the classical wave $\phi(x)$ can create and annihilate particles in quantum theory.

9 Dirac equation

To connect to Einstein's special relativity, for a free particle, we need to change $H = \frac{p^2}{2}$ to

$$\begin{aligned} H &= \sqrt{c^2 + p^2} \\ &= \sqrt{c^2 + (p_1^2 + p_2^2 + p_3^2)} \end{aligned}$$

i.e., $p = (p_1, p_2, p_3)$ in 3D space. If p^2 is small relative to c^2, then a first order Taylor expansion will reduce H to the classical $\frac{p^2}{2}$.

To move to quantum, we again change x and p to X and P with $XP - PX = iI$, e.g., $P = -i(\frac{d}{dx_1}, \frac{d}{dx_2}, \frac{d}{dx_3})$ in the x-basis. But the square root makes things too complicated. So Dirac wants to write

$$\begin{aligned} H &= \sqrt{c^2 + (P_1^2 + P_2^2 + P_3^2)} \\ &= \beta c + \alpha_1 P_1 + \alpha_2 P_2 + \alpha_3 P_3 \end{aligned}$$

This is clearly impossible if p's, β and α's are scalars. But Dirac realized that one can make $(\beta, \alpha_1, \alpha_2, \alpha_3)$ 4×4 matrices, and then one can accomplish the above feat.

This means $h_t(x)$ has four components for each x. This leads to the Dirac equation.

10 What is h_t?

h_t is the universe's bookkeeping of an observer's knowledge or information, i.e., what the observer knows or does not know. If the observer chooses to observe $O = QDQ^*$ at time t, then the universe reveals d_i to the observer according to probability $|\langle h_t, q_i \rangle|^2$. Then the universe changes h_t to q_i which is the embedding of the observed value.