Homework 5: Let's Play LLM

Part 1 : Designing a Task for LLM and Explore the Capability of LLM

Motivation

- Describe the reason why you choose this task :
  When the exist of ChatGpt, many people thought their works will be replaced by AI. Then, how about mathmatician ? Could AI solve math problem ?
- Describe the capability of LLM you want to explore :
  I will test Chatgpt on three domain : computation, high school probability, and proven skill. I think these three domains can measure a human's math ability, so do chatgpt

Task Description

- Describe the task you want to solve :
  The first skill : computation problem will be choose from Calculus, it will derive a function
  The second skill : high school probability will be choose from Combanation math, it will solve Poker problem
  The third skill : proven problem will be choose from Linear algebra, it has to prove Real Spetral Theorem

- Describe the format of the input prompt

  You have to solve three math problem below :

  (1) Consider we are do Calculus homework. if y = (x^3 + 2 * x + 2)^100, find derivative of y

  (2) Consider we are playing a poker game. If a 5-card hand is chosen at random, what is the probability of obataining a flush (all five cards in the hand are in the same suit) ?

  (3) Prove Real Spetral Theorem : here is the theorem,
  Suppose F = R, and T belong to L(V). Then the following are equivalent
  (a) T is self-adjoint

(b) V has an orthonormal basis consisting of eigenvectors of T.
(c) T has a diagonal matrix with respect to some orthonormal basis of V.
Please prove (a), (b), (c) are equivalent

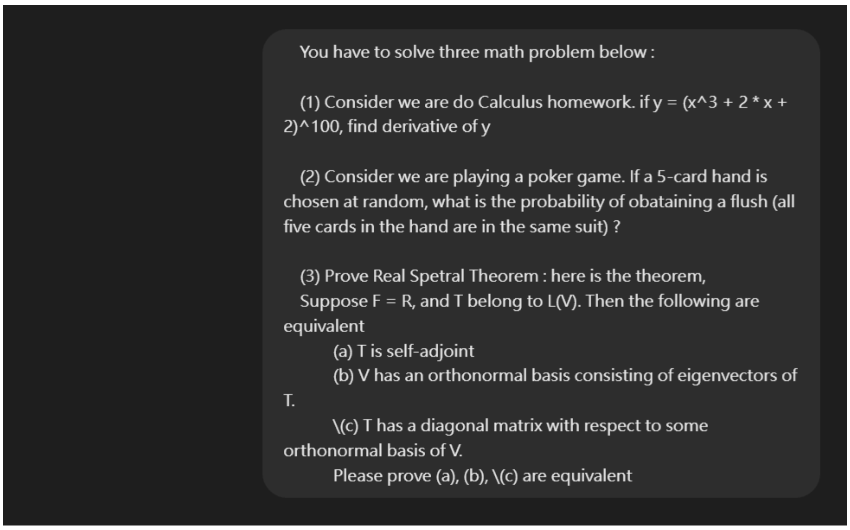
- Describe the ideal output :
  (1) it should use chain rule, thus the derivative of y = 100 * (x^3 + 2 * x + 2)^99 * (3 * x^2 + 2)

  (2) To form the flush, w first decide the suit and then pick 5 cards from the suit. The total number of possible 5 -card flushes is
  $\binom{4}{1} \cdot \binom{13}{5} = \binom{4}{1} \cdot \frac{13!}{5! \cdot 8!} = 4 \cdot 1287 = 5184$
  The probability is 5184/ 2598960 = 0.00198 approximatelly

  (3) First suppose (c) holds, so T has a diagonal matrix with respect to some orthonormal basis of V. A diagonal matrix equals its transpose. Hence T = T*, and thus T is self-adjoint. In other words, (a) holds
  we will prove that (a) implys (b) by indunction on dim V. To get started, note that if dim V = 1, then (a) implys (b). Now assume that dim V > 1 and that (a) implies (b) for all real inner product spaces of smaller dimensoins.
  Suppose (a) holds, so T belong to L(v) is self-adjoint. Let u be an eigenvector of T with L2 norm = 1 . Let U = span(u). Then U is a 1-dimensional subspace of V that is invariant under T. By 7.28 (c), the operator T under orthogonal of U belongs to L(orthogonal of U) is self-adjoint.
  By our induction hypotheisis, there is an orthonormal basis of orthogonal U consisting of eigenvectors of T under orthogonal U. Adjoining u to this orthonormal basis of orthogonal U gives an orthonormal basis of V consisting of eigenvectors of T, completing the proof that (a) implies (b).
  We have proved that (c) implies (a) and that (a) implies (b). Clearly (b) implies (c), completing the proof

Describe your Initial Attempt

- input



You have to solve three math problem below :

(1) Consider we are do Calculus homework. if y = (x^3 + 2 * x + 2)^100, find derivative of y

(2) Consider we are playing a poker game. If a 5-card hand is chosen at random, what is the probability of obataining a flush (all five cards in the hand are in the same suit) ?

(3) Prove Real Spetral Theorem : here is the theorem,
Suppose F = R, and T belong to L(V). Then the following are equivalent
(a) T is self-adjoint
(b) V has an orthonormal basis consisting of eigenvectors of T.
\(c) T has a diagonal matrix with respect to some orthonormal basis of V.
Please prove (a), (b), \(c) are equivalent

- output

problem 1

Let's solve each of the three problems one by one.

Problem 1: Calculus - Finding the Derivative

Given $y = (x^3 + 2x + 2)^{100}$, we need to find $\frac{dy}{dx}$.

To find the derivative, we use the chain rule. Let $u = x^3 + 2x + 2$. Then, $y = u^{100}$.

First, find $\frac{dy}{du}$:
$\frac{dy}{du} = 100u^{99}$

Next, find $\frac{du}{dx}$:
$\frac{du}{dx} = \frac{d}{dx}(x^3 + 2x + 2) = 3x^2 + 2$

Now, apply the chain rule:
$\frac{dy}{dx} = \frac{dy}{du} \cdot \frac{du}{dx}$
$\frac{dy}{dx} = 100u^{99} \cdot (3x^2 + 2)$

Substitute $u = x^3 + 2x + 2$ back into the expression:
$\frac{dy}{dx} = 100(x^3 + 2x + 2)^{99} \cdot (3x^2 + 2)$

So, the derivative is:
$\frac{dy}{dx} = 100(x^3 + 2x + 2)^{99}(3x^2 + 2)$

problem 2

Problem 2: Probability - Obtaining a Flush in Poker

To find the probability of obtaining a flush (all five cards in the hand are of the same suit) in a 5-card poker hand, we proceed as follows:

1. **Total number of 5-card hands:**
   The total number of ways to choose 5 cards out of 52 is given by the combination:
   $\binom{52}{5} = \frac{52!}{5!(52-5)!} = 2,598,960$

2. **Number of ways to get a flush:**

   - There are 4 suits in a deck (hearts, diamonds, clubs, spades).
   - The number of ways to choose 5 cards all from one suit (say hearts) is $\binom{13}{5}$:
     $\binom{13}{5} = \frac{13!}{5!(13-5)!} = 1,287$

   Since there are 4 suits, the total number of flushes is:
   $4 \times 1,287 = 5,148$

3. **Probability of a flush:**
   The probability is the ratio of the number of favorable outcomes to the total number of outcomes:
   $P(\text{flush}) = \frac{\text{Number of flushes}}{\text{Total number of 5-card hands}} = \frac{5,148}{2,598,960}$

   Simplifying this fraction:
   $P(\text{flush}) = \frac{5,148}{2,598,960} \approx 0.00198$

So, the probability of obtaining a flush in a 5-card poker hand is approximately 0.198%.

problem 3

Problem 3: Proof of the Real Spectral Theorem

To prove the Real Spectral Theorem, we need to show the equivalence of the following statements for a linear operator T on a finite-dimensional inner product space V over the real numbers $\mathbb{R}$:

1. T is self-adjoint.
2. V has an orthonormal basis consisting of eigenvectors of T.
3. T has a diagonal matrix with respect to some orthonormal basis of V.

Proof:

(1) $\Longrightarrow$ (2): If T is self-adjoint, then $T = T^*$.

By the spectral theorem for self-adjoint operators in finite dimensions, we know that any self-adjoint operator on a finite-dimensional inner product space has an orthonormal basis consisting of eigenvectors. This directly shows that V has an orthonormal basis consisting of eigenvectors of T.

(2) $\Longrightarrow$ (3): If V has an orthonormal basis consisting of eigenvectors of T.

Let $\{v_1, v_2, \ldots, v_n\}$ be the orthonormal basis consisting of eigenvectors of T. When T is represented in this basis, its matrix representation will be diagonal. Specifically, if $Tv_i = \lambda_i v_i$ for $i = 1, 2, \ldots, n$, then the matrix of T will be:
$\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$

(3) $\Longrightarrow$ (1): If T has a diagonal matrix with respect to some orthonormal basis of V.

Let the orthonormal basis be $\{v_1, v_2, \ldots, v_n\}$. The matrix representation of T in this basis is diagonal, say $\mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$. For any vector $x \in V$, written as $x = \sum_{i=1}^n x_i v_i$:
$T(x) = T\left(\sum_{i=1}^n x_i v_i\right) = \sum_{i=1}^n x_i T(v_i) = \sum_{i=1}^n x_i \lambda_i v_i$

Since T is represented by a diagonal matrix with respect to the orthonormal basis, T is self-adjoint, because the condition $T = T^*$ (where T^* is the adjoint of T) is satisfied by the diagonal matrix.

Therefore, the three statements are equivalent, proving the Real Spectral Theorem.

- <u>Analysis</u>
  For problem 1,2, it do pretty well. The answer of it is correct.
  While in problem 3, it doesn't do well on (1) => (2) subproblem. It use the spectral theorem result to prove spectral theorem itself, which is not allowed in proven.

Describe Improvements

We will address the problem 3 proven by provided more helpping lemma and definition

- <u>I will give the following lemma</u>
  lemma : self-adjoint operators and invariant subspaces
  Suppose T belong to inner product space of V and U is a subspace of V that is invaraint under T. Then,
  (a) orthogonal of U is invariant under T
  (b) T under U is self-adjoint
  (c) T under orthogonal of U is self-adjoint

let's try it again

- <u>input</u>

Prove Real Spetral Theorem : here is the theorem,
Suppose F = R, and T belong to L(V). Then the following are equivalent
(a) T is self-adjoint
(b) V has an orthonormal basis consisting of eigenvectors of T.
\(c) T has a diagonal matrix with respect to some orthonormal basis of V.
Please prove (a), (b), \(c) are equivalent

here is the helpping lemma:
lemma : self-adjoint operators and invariant subspaces
Suppose T belong to inner product space of V and U is a subspace of V that is invaraint under T. Then,
(a) orthogonal of U is invariant under T
(b) T under U is self-adjoint
\(c) T under orthogonal of U is self-adjoint

- output

To prove the Real Spectral Theorem, we need to show the equivalence of the following three statements for a linear operator T on a finite-dimensional inner product space V over $\mathbb{R}$:

1. T is self-adjoint.
2. V has an orthonormal basis consisting of eigenvectors of T.
3. T has a diagonal matrix with respect to some orthonormal basis of V.

We'll show that (a) ⇒ (b), (b) ⇒ (c), and (c) ⇒ (a).

(a) ⇒ (b):

1. **Assume T is self-adjoint:** $T = T^*$, meaning $\langle Tv, w\rangle = \langle v, Tw\rangle$ for all $v, w \in V$.
2. **Eigenvalue Existence:** Since T is self-adjoint, all its eigenvalues are real. To see why, consider the characteristic polynomial of T, which has real coefficients because T is a real operator.
3. **Invariant Subspace:** Let λ be an eigenvalue of T with eigenvector v. Then the subspace $\langle v\rangle$ spanned by v is invariant under T. By the lemma, the orthogonal complement $\langle v\rangle^\perp$ is also invariant under T.
4. **Inductive Step:** Consider T restricted to $\langle v\rangle^\perp$. This restriction is also self-adjoint. By induction on the dimension of V, T restricted to $\langle v\rangle^\perp$ has an orthonormal basis of eigenvectors. Combining v and the orthonormal basis of eigenvectors of the restriction gives us an orthonormal basis for V consisting of eigenvectors of T.

(b) ⇒ (c):

1. **Orthonormal Basis of Eigenvectors:** Assume V has an orthonormal basis $\{e_1, e_2, \ldots, e_n\}$ consisting of eigenvectors of T.
2. **Diagonal Representation:** In this basis, T is diagonal. Specifically, if $Te_i = \lambda_i e_i$, then T has a diagonal matrix representation with the eigenvalues λ_i on the diagonal. This is because $T(e_j) = \sum_{i=1}^n a_{ij}e_i$ and orthonormality implies $a_{ij} = 0$ for $i \neq j$.

(c) ⇒ (a):

1. **Diagonal Matrix:** Assume T has a diagonal matrix with respect to some orthonormal basis $\{e_1, e_2, \ldots, e_n\}$.
2. **Self-Adjoint Property:** If T is diagonal with respect to this orthonormal basis, T is self-adjoint. This is because in this basis, $T = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ where λ_i are real (since T is real). The adjoint of a diagonal matrix with real entries is itself, hence $T = T^*$.

Conclusion:

Since we have shown (a) ⇒ (b), (b) ⇒ (c), and (c) ⇒ (a), the three statements (a), (b), and (c) are equivalent. Thus, the Real Spectral Theorem is proven.

- Analysis
By providing helping lemmas, it no longer prove Spectral Theorem by Spectral Theorem. Instead, it use the provided lemma to further prove the result.

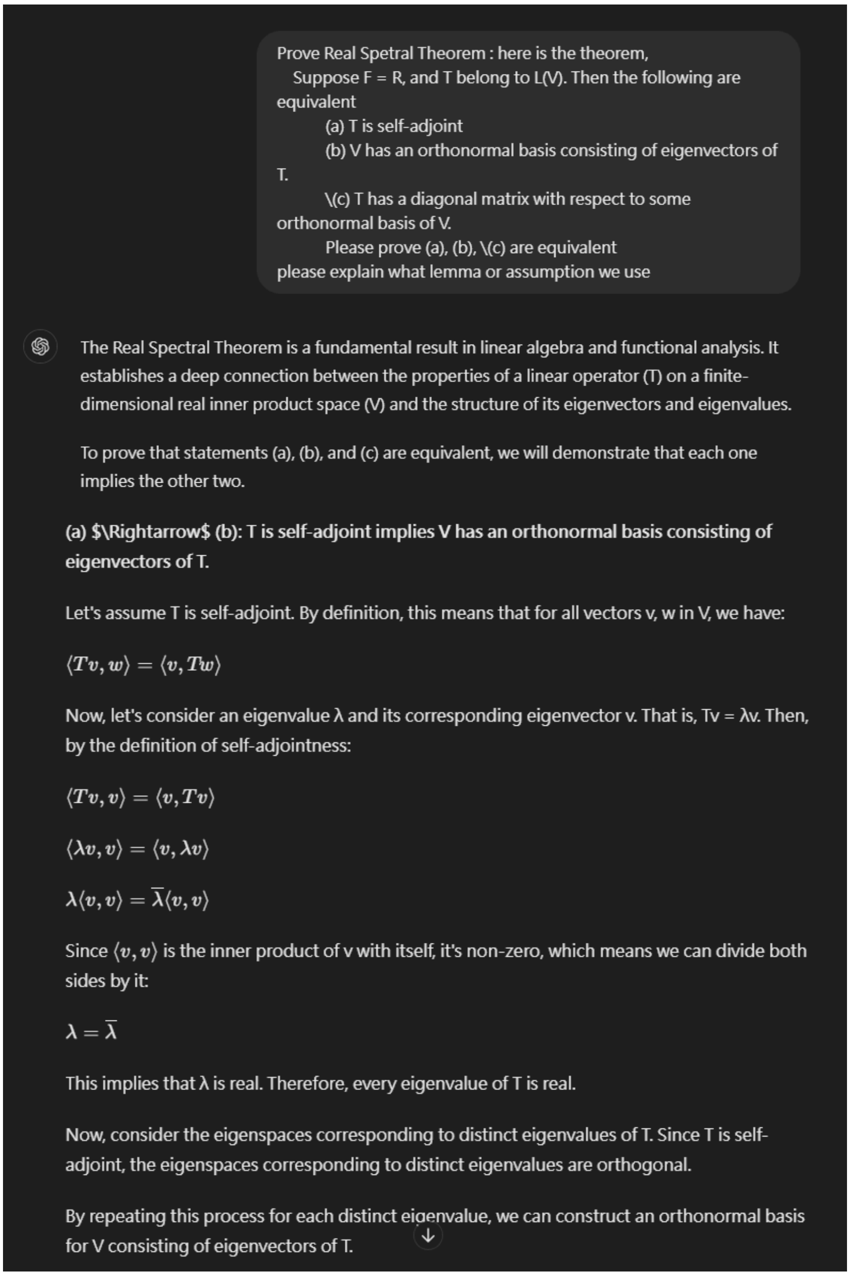
Part 2: Questions Answering

Question 1

1. Besides providing a more detailed description of the question in the prompt,what other methods can be employed to improve the accuracy of the responses?
   We could ask LLM to elaborate what it think about when responding.　The mothod is related to chain of thought prompting

2. Please provide an example demonstrating the impact of implementing the aforementioned methods on a prompt.

   Compared to the Part1. Initial Attempt, it doesn't prove Spectral theorem by Spectral theorem is this time, which is a improvement



Prove Real Spetral Theorem : here is the theorem,
Suppose F = R, and T belong to L(V). Then the following are equivalent
(a) T is self-adjoint
(b) V has an orthonormal basis consisting of eigenvectors of T.
\(c) T has a diagonal matrix with respect to some orthonormal basis of V.
Please prove (a), (b), \(c) are equivalent
please explain what lemma or assumption we use

The Real Spectral Theorem is a fundamental result in linear algebra and functional analysis. It establishes a deep connection between the properties of a linear operator (T) on a finite-dimensional real inner product space (V) and the structure of its eigenvectors and eigenvalues.

To prove that statements (a), (b), and (c) are equivalent, we will demonstrate that each one implies the other two.

(a) $\Rightarrow$ (b): T is self-adjoint implies V has an orthonormal basis consisting of eigenvectors of T.

Let's assume T is self-adjoint. By definition, this means that for all vectors v, w in V, we have:

$\langle Tv, w\rangle = \langle v, Tw\rangle$

Now, let's consider an eigenvalue λ and its corresponding eigenvector v. That is, Tv = λv. Then, by the definition of self-adjointness:

$\langle Tv, v\rangle = \langle v, Tv\rangle$

$\langle \lambda v, v\rangle = \langle v, \lambda v\rangle$

$\lambda\langle v, v\rangle = \overline{\lambda}\langle v, v\rangle$

Since $\langle v, v\rangle$ is the inner product of v with itself, it's non-zero, which means we can divide both sides by it:

$\lambda = \overline{\lambda}$

This implies that λ is real. Therefore, every eigenvalue of T is real.

Now, consider the eigenspaces corresponding to distinct eigenvalues of T. Since T is self-adjoint, the eigenspaces corresponding to distinct eigenvalues are orthogonal.

By repeating this process for each distinct eigenvalue, we can construct an orthonormal basis for V consisting of eigenvectors of T.

Question 2

- Question: How to make the output of LLM safer without fine-tune the original parameter
- Motivation: Human find that chatgpt may provide some dangerous informatoin like how to make TNT. However, fine-tuning the original parameter is resource-consuming because LLM has a huge parameter size. For example GPT3 has 175 billion parameters, it may takes lots of GPU to fine-tune such things
- Challenge
  - How could we use the original LLM output
  - How could we let output of LLM safer
  - How could we not affect the original parameters
- Solution
  we could make a wrapper on the output of LLM. Assume LLM acquire all the information when pretraining, now we have to do is to align model output to the right format.
  It could be a model-based RL
- Reference
  LIMA: Less Is More for Alignment (https://arxiv.org/abs/2305.11206)
  RAIN: Your Language Models Can Align Themselves without Finetuning (https://arxiv.org/abs/2309.07124)