Examination questions

1. Say you have a database a movie ratings. That is you have data on different users our website and the ratings that they have given for movies.

How would you recommend new movies to the users of your website?

Walk us through the steps Bob this algorithm.

Explain all the assumptions you are making. For example what kind of data would you collect?

1. Say you wanted to use the principle components for a regression problem. Why would you want to do this? How would you go about doing this?
2. The results rest on a key tension between validity only generate valid strings and breadth generate many strings

Things are on a spectrum from hallucination which is a failure of validity to mode collapse which is a failure of breadth

My interpretation is that these are true of all learning systems that is given a machine learning algorithm capital L there will always be a tradeoff between validity only generate valid strings and breadth generate many strings from a language capital K

Here’s a draft for an assignment question based on your points:

**Assignment Question:**  
Discuss the inherent trade-offs in machine learning systems between validity and breadth in language generation tasks. Using examples from the abstraction and reasoning corpus (ARC) and large language models (LLMs), analyze how the spectrum from hallucination (failure of validity) to mode collapse (failure of breadth) manifests. Evaluate whether it is possible to design a learning algorithm, LL, that optimally balances the generation of valid strings while maintaining broad generalization across a language KK. Support your discussion with theoretical and practical insights.

Let me know if you’d like any refinements or additions!

4. how would you do dimensionality reduction of text. Outline steps. Encoding. One hots.

5. what is linear dimensionality reduction

6. how would you d non-linear dimensionality reduction