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.

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

5. what is linear dimensionality reduction

6. how would you do non-linear dimensionality reduction

7. Sure, I'd be happy to help! Here are some theoretical assignment questions on unsupervised machine learning:

1. **Clustering Concepts**: Explain the difference between hierarchical clustering and partitional clustering. Provide examples of scenarios where each method might be preferred.
2. **Dimensionality Reduction**: Discuss the importance of dimensionality reduction in machine learning. Compare and contrast PCA (Principal Component Analysis) with t-SNE (t-Distributed Stochastic Neighbour Embedding) in terms of their methodology and applications.
3. **Applications of Clustering**: Describe how clustering algorithms can be used for anomaly detection. What are some challenges associated with using clustering for this purpose?
4. **Evaluation Metrics**: What are the different metrics used to evaluate the performance of clustering algorithms? Explain the advantages and limitations of each.
5. **Ethical Considerations**: Discuss the ethical implications of using unsupervised learning methods, particularly in contexts like customer segmentation or social profiling. How can biases arise, and what measures can be taken to mitigate them?

7) The principles of dimensionality reduction are also used in modern machine learning techniques like large language models. Refer to the diagram below and suggest how dimensionality reduction can be used in situations where there is massive amounts of data.

Given a huge amount of data, say in the context of electronic healthcare records, how would you use dimensionality reduction with a supervised machine learning method?

8) There's a question on supervised principal components analysis or semi supervised machine learning methods. What i'm looking for is a new on stun the standing OB how is the PCA principle components analysis or an auto encoder can be used in conjunction with random forest or a logistic regression algorithm.

How can this be used in conjunction with a supervised machine learning algorithm?

What I'm looking for is a nuanced understanding of how dimensionality reduction techniques can be used to reduce the input features and then use these reduced input features as inputs to a supervised machine learning algorithm.

9) what are some ways to create nonlinear dimensionality reduction techniques? For example, principal components analysis is a linear dimensionality reduction technique. Think back about all the techniques that we have studied in both the unsupervised and supervised machine learning classes. Now think about way to have create nonlinear dimensionality reduction technique.

10) Richard feynman question on a heuristic based algorithm to solve a certain task. For example if a heuristic is used repeatedly then you keep that heuristic on the top and then that becomes the topmost heuristic or the 1st heuristic that you used next time you see a problem. Ask the students to come up with a heuristic based algorithm to solve a certain task. Ask them to write seed or code. Ask them to run a sample run of this particular algorithm on sample problem.

<https://www.youtube.com/watch?v=ipRvjS7q1DI>

11) question on using correlation based distance for hierarchical clustering to find shoppers with similar tastes in shopping data. This can also be applied to movie data. For example, the same problem of classifying movie users or clustering movie users can be applied using hard impute techniques like principle components analysis or using clustering techniques like hierarchical clustering with correlation based distance.

10)

Image below

A screen shot of a computer

Description automatically generated