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) **A Classic IMO Problem**

Find all pairs of integers (x,y)(x, y)(x,y) such that:

x^2 + y^2 = 1 + 2xy

8) bayes principle

You have two envelopes, one containing twice as much money as the other. You pick one envelope at random. Before opening it, you are told that you can switch envelopes if you'd like. Would switching always increase your expected gain? (Ask students to analyze and justify.)

9) Three friends go to a restaurant and order food worth $30. They each contribute $10. Later, the manager realizes there was an error in billing and the actual cost is $25. He gives $5 to the waiter to return to the friends. The waiter, however, pockets $2 and gives $1 back to each friend. Now, each friend has paid $9, which totals $27, and the waiter kept $2, making $29. Where is the missing dollar?

How would a computer solve it? Walk through the steps, the assumptions you are making and how you would design a computer algorithm to solve this.

10) function cv(k = 10, ..)

What is k

How should you choose k?

k = 13

cv(k = 10,..)

print(k)

cv(k <- 10, …)

print(k)

what is the value of k

practice questions

7) question on interpreting principle component loadings and principle component scores based on either a plot or a table like in the book table 12.1 on page 503 of introduction to statistical learning in r and a figure like the biplot table and figure on page 502 of introduction to statistical learning in r.

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 are 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?

Another question on how word embeddings are used and how these are similar to PCA and how these can be used in large language models. We had some slides that were covered in teaching in this class.

<https://youtu.be/wjZofJX0v4M?si=TU4pAiDMolliTmpZ&t=837>

Also another question. Suppose you have been given audio somehow arasta derive word embedding of the following form shown below. That is, you're given thousands of different words in the English language and you are asked to make a more compact representation using numbers of the form shown below.

How would you go about doing this?

Image below

A screenshot of a computer

Description automatically generated

8) There's a question on supervised principal components analysis or semi supervised machine learning methods. What I am looking for is an understanding of how PCA principal 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) please write an algorithm for hierarchical clustering. You can write in pseudo code.

Then please explain this algorithm diagrammatically using an example. Give the example of data. You can use two-dimensional data and then plot the resulting dendrogram.

11) say at your first job or your first project you're given access to some shopping data. Your goal is to find shoppers who have similar tastes. How would you develop an algorithm to do this?

Explain all the steps.

Layout all your assumptions. Explain how you would go about analysing the data. Explain what kind of data you're dealing with and what format it is in. Then explain how you would go about using some techniques to solve this problem.

Nuances - if shopped then 1, if never shopped then 0

Or use volumes of purchases (continuous)

11) supervised question on gini index

See image below

A screenshot of a whiteboard

Description automatically generated

12) 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.

This question can also be posed by showing data like the one shown below and asking what kind of clustering with what kinds of parameters (dissimilarity metric, linkage etc) to use.

User behaviour data shown below.

Image below.

A graph of different colored lines

Description automatically generated

12) how to reorder a dendrogram and how many reorderings are there?

13) ARC problem. Assume you have been given the following image. How would you use a machine learning algorithm to solve it? What kind of machine learning algorithm would you use? Walk us through the steps of designing this algorithm.

Image below.

A screenshot of a computer program

Description automatically generated

12) this is an image of all the micro impacts on the surface of a space shuttle. This shows all data before any of the challenger and Columbia accidents. Based on this image can you suggest a way of how to provide better shielding to the space shuttle? Are there any particular things that we should be aware of before we do any analysis?

Image below

A diagram of a space shuttle

Description automatically generated

13)

Image below

A screen shot of a computer

Description automatically generated