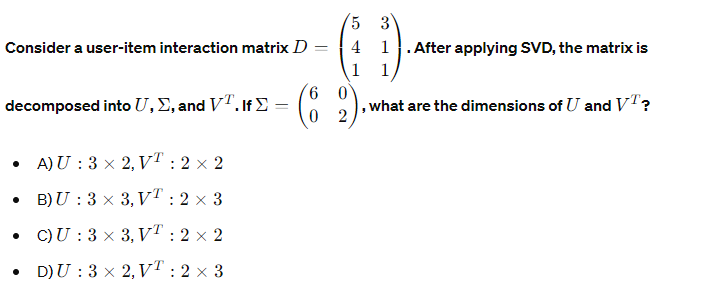
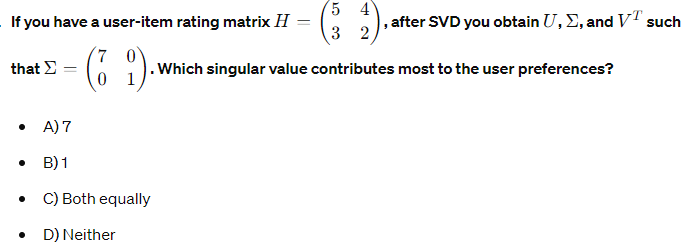
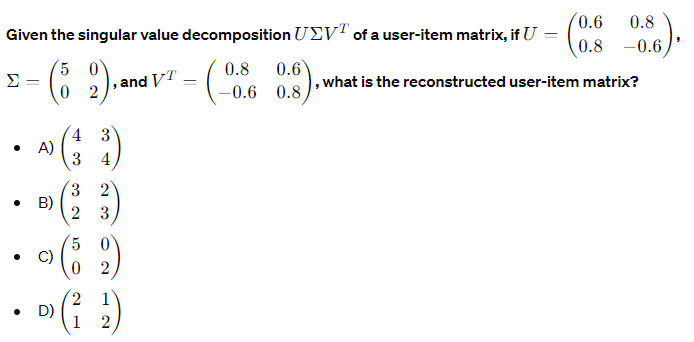
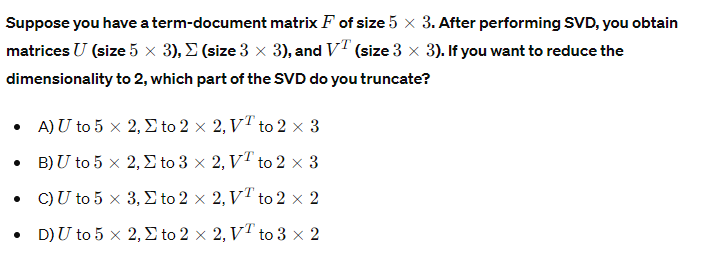
**Natural Language Processing (Final Exam - 2024)**

1. Which of the following is NOT a direct application of SVD in data analysis?\*\*
   1. Image compression
   2. Noise reduction
   3. Matrix inversion
   4. Data encryption
2. In SVD, a matrix A is decomposed into three matrices U, Sigma, and VT. What does the matrix Sigma represent?
   1. The original matrix A
   2. The left singular vectors
   3. The singular values
   4. The right singular vectors
3. How can SVD be used to identify the most important features in a dataset?\*\*
   1. By analyzing the matrix U
   2. By looking at the largest singular values in Sigma
   3. By transforming the dataset using VT
   4. By reconstructing the dataset with all singular values
4. What role does SVD play in the context of a recommender system?
   1. It clusters users into different groups
   2. It predicts user preferences by decomposing the user-item interaction matrix
   3. It creates new items based on user preferences
   4. It increases the number of recommendations a user receives
5. Which matrix in the SVD decomposition of a user-movie interaction matrix represents the latent factors for items?
   1. U
   2. Sigma
   3. VT
   4. None of the above
6. How does SVD improve the performance of a recommender system?
   1. By directly using user ratings
   2. By approximating the user-item interaction matrix to predict missing values
   3. By ignoring user-item interactions
   4. By increasing the dimensionality of the dataset
7. In which scenario would you most likely apply SVD in both data analysis and a recommender system?
   1. When dealing with large-scale text data for sentiment analysis
   2. When analyzing social network graphs for community detection
   3. When performing collaborative analysis on a user-item matrix to recommend movies
   4. When compressing audio files for better storage
8. Which of the following steps is necessary to apply SVD for a recommender system?
   1. Normalizing the user-item matrix
   2. Adding random noise to the matrix
   3. Removing the largest singular values
   4. Transposing the user matrix
9. Which of the following best describes the "truncated SVD"?
   1. Using only a subset of the largest singular values and corresponding vectors
   2. Discarding all singular values and vectors
   3. Applying SVD without normalization
   4. Using SVD on a transposed matrix
10. How can SVD be integrated with neural networks to enhance the performance of recommender systems?
    1. By using SVD to preprocess the data before feeding it into the network
    2. By replacing hidden layers with SVD components
    3. By using SVD to post-process the neural network output
    4. By combining the output of SVD with the input features of the network
11. In the context of Natural Language Processing (NLP), how is SVD used for Latent Semantic Analysis (LSA)?
    1. To reduce the dimensionality of term-document matrices and uncover hidden structures in the data
    2. To increase the number of terms in the vocabulary
    3. To directly translate documents into multiple languages
    4. To generate new terms based on the existing vocabulary
12. 1



1. In a typical Seq2Seq model using an encoder-decoder architecture, what does the encoder do?
   1. Generates output sequences
   2. Encodes the input sequence into a fixed-length context vector
   3. Translates the input sequence into another language
   4. Decodes the context vector into the output sequence
2. Which of the following techniques is often used to improve the performance of Seq2Seq models for long sequences?
   1. Data augmentation
   2. Attention mechanism
   3. Dropout regularization
   4. Batch normalization
3. In the context of Seq2Seq models, what is the main advantage of using an attention mechanism?
   1. Reduces the number of parameters
   2. Allows the model to focus on different parts of the input sequence when generating each part of the output sequence
   3. Eliminates the need for training data
   4. Speeds up the training process
4. What is a key difference between traditional Seq2Seq models with RNNs and transformer models?
   1. Transformers do not use convolutional layers
   2. Transformers rely on self-attention mechanisms rather than recurrence
   3. Transformers can only handle fixed-length sequences
   4. Transformers are used only for image processing tasks
5. Which component in the transformer architecture is responsible for capturing dependencies between all elements of the input sequence simultaneously?
   1. Convolutional layer
   2. Recurrent layer
   3. Self-attention layer
   4. Pooling layer
6. In the transformer architecture, what is the purpose of the position encoding?
   1. To encode the position of tokens in the sequence since the model lacks recurrence
   2. To initialize the weights of the model
   3. To normalize the input data
   4. To reduce the dimensionality of the input data
7. Which transformer model has become the foundation for many state-of-the-art natural language processing tasks?
   1. BERT
   2. AlexNet
   3. VGG
   4. ResNet
8. Which application is NOT typically associated with Seq2Seq models or transformers?\*\*
   1. Language translation
   2. Text summarization
   3. Image classification
9. How do transformers handle long-range dependencies in sequences more effectively than traditional RNNs?
   1. By using convolutional layers
   2. By employing a hierarchical structure
   3. By applying self-attention mechanisms
   4. By reducing the sequence length through pooling layers
10. In machine translation using transformers, what role does the encoder play?
    1. Encodes the target language sequence
    2. Generates the translated output sequence
    3. Encodes the source language sequence into contextual embeddings
    4. Directly maps input sequences to output sequences without transformation
11. What is one advantage of using the BERT model over traditional Seq2Seq models for natural language understanding tasks?\*\*
    1. BERT uses bidirectional context, allowing it to understand the context from both directions
    2. BERT requires less training data
    3. BERT can process images and text simultaneously
    4. BERT has fewer parameters, making it faster to train
12. In a transformer-based text generation model, what is the purpose of using masked self-attention in the decoder?
    1. To prevent the model from attending to irrelevant parts of the input sequence
    2. To ensure the model only attends to previously generated tokens and not future tokens
    3. To reduce the computational complexity of the model
    4. To improve the model's ability to handle long sequences
13. For a Seq2Seq model with attention mechanism applied to text summarization, how does the attention mechanism benefit the summarization task?
    1. It reduces the training time required for the model
    2. It allows the model to focus on the most relevant parts of the input text when generating the summary
    3. It increases the overall length of the generated summary
    4. It simplifies the architecture of the model
14. How is the transformer model adapted for the task of question answering in models like BERT?
    1. By adding a recurrent layer to handle sequences
    2. By fine-tuning the pre-trained model on a question-answering dataset
    3. By using transformers only in the encoder part of the model
    4. By combining transformers with convolutional neural networks
15. What is a Large Language Model (LLM)?
    1. A model that primarily processes image data
    2. A model with a large number of parameters designed to understand and generate human language
    3. A model used for financial forecasting
    4. A model specifically for translating programming languages
16. Which of the following is an example of a Large Language Model?
    1. ResNet
    2. GPT-3
    3. Inception
    4. YOLO
17. What does "GPT" in GPT-3 stand for?
    1. General Processing Transformer
    2. Graphical Pre-trained Transformer
    3. Generative Pre-trained Transformer
    4. General Purpose Translator
18. How do Large Language Models typically get trained?
    1. By using supervised learning with labeled data only
    2. By using unsupervised learning with large text corpora
    3. By using reinforcement learning exclusively
    4. By training on image datasets
19. What is a common application of Large Language Models like GPT-3?
    1. Diagnosing medical images
    2. Generating human-like text for chatbots
    3. Recognizing objects in images
    4. Predicting stock market trends
20. What does the term "pre-training" refer to in the context of LLMs?
    1. Training the model on specific, labeled datasets before fine-tuning
    2. Training the model on a large, diverse dataset before fine-tuning on a specific task
    3. Training the model using only unsupervised methods
    4. Initializing the model weights randomly before training
21. Which of the following challenges is associated with Large Language Models?
    1. They can easily be run on any consumer-grade hardware
    2. They always produce accurate and unbiased results
    3. They require significant computational resources and data for training
    4. They are easy to interpret and understand
22. What is the primary architecture used in most Large Language Models today?
    1. Recurrent Neural Networks (RNNs)
    2. Convolutional Neural Networks (CNNs)
    3. Transformer networks
    4. Support Vector Machines (SVMs)
23. Which Large Language Model is known for its bidirectional context understanding?
    1. BERT
    2. GPT-2
    3. ELMo
    4. Transformer-XL
24. How do LLMs like GPT-3 generate text?
    1. By predicting the next word in a sequence based on the context of previous words
    2. By matching patterns in image data
    3. By using predefined templates for text generation
    4. By querying a database of pre-written texts
25. What is the main difference between GPT-2 and GPT-3?
    1. GPT-2 uses a different architecture than GPT-3
    2. GPT-3 has significantly more parameters and training data compared to GPT-2
    3. GPT-2 is designed for image generation, while GPT-3 is for text generation
    4. GPT-2 is a bidirectional model, while GPT-3 is unidirectional
26. Why is the ability to handle long-range dependencies important for LLMs?
    1. It allows the model to process images more effectively
    2. It enables the model to understand context over long paragraphs of text
    3. It improves the model's performance on short text sequences
    4. It reduces the computational cost of training the model
27. What does the term "context window" refer to in LLMs?
    1. The maximum length of text the model can consider at once
    2. The size of the training dataset
    3. The number of layers in the model
    4. The amount of computational resources needed for training
28. What is the primary goal of fine-tuning a pretrained LLM?
    1. To train the model from scratch on a new dataset
    2. To adjust the pretrained model to perform well on a specific downstream task
    3. To reduce the model size for deployment
    4. To convert the model to handle image data
29. When fine-tuning a pretrained LLM, what does freezing layers mean?
    1. Training all layers of the model equally
    2. Only updating the weights of certain layers while keeping the others constant
    3. Increasing the learning rate for certain layers
    4. Reducing the size of the model
30. Which of the following is a common strategy to prevent overfitting during fine-tuning?
    1. Using a very high learning rate
    2. Increasing the number of parameters in the model
    3. Implementing techniques such as dropout and early stopping
    4. Avoiding the use of validation datasets
31. In fine-tuning an LLM for a classification task, which loss function is commonly used?
    1. Mean Squared Error (MSE)
    2. Cross-Entropy Loss
    3. Hinge Loss
    4. Mean Absolute Error (MAE)
32. What is the purpose of using a validation set during the fine-tuning process?
    1. To increase the training data size
    2. To monitor the model’s performance on unseen data and prevent overfitting
    3. To determine the final test accuracy
    4. To fine-tune the hyperparameters of the model
33. During fine-tuning, why might you choose to use a lower learning rate compared to the pretraining phase?\*\*
    1. To prevent drastic changes in the pretrained weights
    2. To speed up the training process
    3. To avoid using too much computational power
    4. To ensure the model forgets the pretrained knowledge
34. What is transfer learning in the context of fine-tuning LLMs?
    1. Training a new model from scratch using the same architecture
    2. Transferring the knowledge from a pretrained model to a new task
    3. Converting a model to handle different types of data
    4. Using the model for multiple tasks simultaneously without any modification
35. In the context of fine-tuning LLMs, what is meant by the term "domain adaptation"?
    1. Adapting the model to perform in a completely different modality (e.g., from text to image)
    2. Adjusting the pretrained model to work effectively on data from a specific domain (e.g., medical texts)
    3. Changing the model architecture to handle different tasks
    4. Modifying the input data format to match the model requirements
36. What is the purpose of using data augmentation during the fine-tuning of LLMs?
    1. To increase the diversity of the training data and prevent overfitting
    2. To reduce the size of the training dataset
    3. To simplify the model architecture