Combined Full Questions and Answers (Chapters 3 to 13)

Chapter 3-4: Statistical Language Models, Word2Vec, GloVe, TF-IDF, Bias

True/False:

Q1. Lemmatization is more computationally expensive than stemming.

Answer: True

Q2. Sigmoid function output is between -1 and 1.

Answer: False

Q3. SentencePiece requires tokenized input data.

Answer: True

Q4. Scalar multiplication is a non-linear operation.

Answer: False

Q5. Dependency parsing mainly focuses on word order.

Answer: False

Multiple Choice:

Q6. In dependency parsing, the head word is called:

(a) Prime (b) Root (c) Parent (d) Core

Answer: (a) Prime

Q7. Study of internal word structure is:

(a) Syntax (b) Phonology (c) Morphology (d) Semantics

Answer: (c) Morphology

Q8. In gradient descent, what is updated?

(a) Parameters (b) Loss (c) Features (d) Epochs

Answer: (a) Parameters

Q9. Dependency parsing reveals:

(a) Sentence semantics (b) Linear sentence form (c) Alphabetical order (d) Morphological structure

Answer: (d) Morphological

Chapter 5: Neural Language Models (RNN, LSTM, Attention)

True/False:

Q10. CNN filters act like n-gram detectors in NLP.

Answer: True

Q11. Padding helps uniform sentence length handling.

Answer: True

Q12. LSTMs use gates to manage flow of information.

Answer: True

Q13. Attention helps model focus on important input parts.

Answer: True

Q14. Fixed vector is the only output in attention models.

Answer: False

Multiple Choice:

Q15. Which combats vanishing gradients?

(a) CNN (b) RNN (c) LSTM (d) Vanilla NN

Answer: (c) LSTM

Q16. Hierarchical convolution helps in:

(a) Speed (b) Complexity handling (c) Activation (d) Optimizing pooling

Answer: (b) Complexity handling

Chapter 6: Transformers (Self-Attention, RoPE)

True/False:

Q17. Feed-forward networks introduce non-linearity.

Answer: True

Q18. Multi-query attention reduces computation.

Answer: True

Q19. Sliding window attention attends locally.

Answer: True

Q20. Rotary embeddings inject position info.

Answer: True

Q21. Cross-attention relates decoder to encoder.

Answer: True

Multiple Choice:

Q22. NOT a benefit of multi-head attention:

(a) Multiple perspectives (b) Enhanced parallelism (c) More expressive (d) Diverse focus

Answer: (b) Enhanced parallelism

Q23. Purpose of residual connections:

(a) Adding more parameters (b) Solve vanishing gradient (c) Replace attention (d) Increase loss

Answer: (b) Solve vanishing gradient

Q24. Why masking in decoder attention:

(a) Attend all tokens (b) Prevent future info (c) Copy past tokens (d) Speed up decoding

Answer: (b) Prevent future info

Q25. Advantage of RoPE:

(a) Better relative position capture (b) Faster computation (c) Larger models (d) Improved tokenization

Answer: (a) Better relative position capture

Q26. Role of Query-Key:

(a) Find similarity (b) Project embeddings (c) Determine attention (d) Form input tokens

Answer: (c) Determine attention

(Continuing in next batch due to long length...)

Chapter 7: Language Model Pretraining (BERT, T5, GPT)

True/False:

Q27. Encoder-only models use Causal Language Modeling.

Answer: False

Q28. The [SEP] token is used to separate sentences.

Answer: True

Q29. Decoder-only models have separate encoders.

Answer: False

Q30. BERT uses masked tokens prediction.

Answer: True

Q31. T5 uses span-based corruption for pretraining.

Answer: True

Multiple Choice:

Q32. Special tokens used in BERT for sentence pairs:

(a) [PAD], [MASK] (b) [SEP], [CLS] (c) [UNK], [SEP] (d) [CLS], [PAD]

Answer: (b) [SEP], [CLS]

Q33. Pretraining objective for T5:

(a) Next Token Prediction (b) Translation (c) Span Corruption (d) MLM

Answer: (d) Span Corruption

Q34. BERT NSP task determines:

(a) Sentiment (b) Next Sentence Link (c) Word similarity (d) Masked tokens

Answer: (b) Next Sentence Link

Chapter 8: Fine-Tuning and Alignment (RLHF, DPO, CoT, ToT, GoT)

True/False:

Q35. CoT enables multiple reasoning paths.

Answer: True

Q36. ToT framework allows parallel path exploration.

Answer: True

Q37. Prompt engineering modifies model weights.

Answer: False

Q38. GoT forms clear sequential paths.

Answer: False

Q39. ToT uses DFS and BFS for problem solving.

Answer: True

Multiple Choice:

Q40. Goal of prompt engineering:

(a) Design effective prompts (b) Fine-tune LLMs (c) Build loss function (d) Tokenize input

Answer: (a) Design effective prompts

Q41. In-context learning implies:

(a) Model retraining (b) Learning from examples in prompts (c) RLHF application (d) Labeling

Answer: (b) Learning from examples in prompts

Q42. Example of prompting:

(a) Summarize this text (b) Mask this token (c) Fine-tune model (d) Optimize decoder

Answer: (a) Summarize this text

Q43. How in-context learning helps:

(a) Regularize model (b) Avoid retraining (c) Clean data (d) Pretrain

Answer: (b) Avoid retraining

Q44. CoT advantage:

(a) Faster training (b) Multi-step reasoning (c) Shorter inputs (d) More loss

Answer: (b) Multi-step reasoning

(Continuing next set for Chapters 9-13...)

Chapter 9: Efficient Fine-Tuning (LoRA, Quantization, BitFit)

True/False:

Q45. Pruning removes whole layers.

Answer: False

Q46. Mixed precision uses only FP32.

Answer: False

Q47. KL divergence used in distillation.

Answer: True

Q48. LoRA and BitFit help fine-tune large models.

Answer: True

Q49. Black-box knowledge distillation is possible.

Answer: True

Multiple Choice:

Q50. Hyperparameter balancing distillation and task loss:

(a) Beta (b) Alpha (c) Delta (d) Gamma

Answer: (b) Alpha

Q51. Soft labels refer to:

(a) Labels from fine-tuned model (b) Raw logits (c) Ground truth (d) Attention weights

Answer: (c) Ground truth

Q52. Benefit of quantisation:

(a) Faster training (b) Smaller model size (c) More parameters (d) Slower memory

Answer: (b) Smaller model size

Q53. Which is NOT a prompt compression technique:

(a) Context dropout (b) Prefix tuning (c) Keyword extraction (d) Prompt distillation

Answer: (c) Keyword extraction

Q54. Mixed precision achieves:

(a) Balancing FP16 and FP32 (b) Only FP32 (c) Only FP16 (d) Slower inference

Answer: (a) Balancing FP16 and FP32

Chapter 10: Augmented Large Language Models (RAG)

(True/False and MCQs will continue similarly...)