Q65. Explain overfitting & regularization.

Q66. Explain bias-variance tradeoff?

Q67. How to handle data imbalance issues?

Q68. Explain Gradient descent & stochastic gradient descent. which one would you prefer?

Q69. Can you explain logistic regression & can you derive gradient descent for logistic regression.

Q70. what do eigenvalues & eigenvectors mean in PCA?

eigenvector is the direction of the hyperplane where data is more spread.

eigenvalues how dataset is spread out on the line which is eigen vector.

Q71. Explain different types of optimizers — How is Adam optimizer different from Rmsprop?

Q72. what are different types of activation functions & explain about vanishing gradient problem.

Q73. what do L1 & L2 regularization mean and when would you use L1 vs L2? Can you use both?

⇒ L1 regularization can address the multicollinearity problem by constraining the coefficient norm & pinning some coefficient values to 0.

L1 regularization can be a feature selection.

⇒ Loss if outliers is more in L2 than L1 because L2 uses square of the weights, whereas L1 uses absolute values of the weights. So cost only increases linearly.

Q14. Can you use MSE for evaluating your classification problem instead of cross entropy?

Assume, 6 class classification problem.

True probabilities = [1, 0, 0, 0, 0, 0]

Case 1: predicted prob: [0.2, 0.16, 0.16, 0.16, 0.16, 0.16]

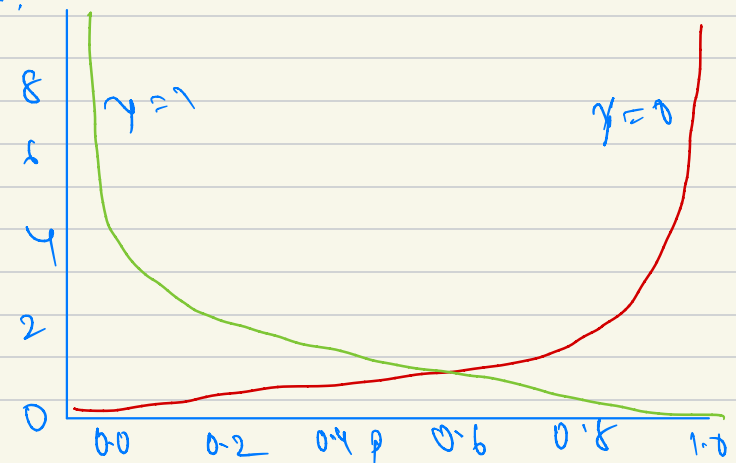
Case 2: " " : [0.4, 0.5, 0.1, 0, 0, 0]

MSE for case 1 = 0.128 → predicting correctly but loss is higher than case 2.

MSE for case 2 = 0.1033

Q15. How does loss curve for cross entropy look?

If we consider $y=0$ as the expected output, the loss is less for small values, and as the p (predicted value of y by algorithm) increases the loss increases rapidly.



Q76. what does the "minus" means in cross entropy?

Consider cross entropy

$$-y_i \log p_i$$

We know $0 \le p_i \le 1$

if $p_i \le 1$ $\ln p_i \le 0$

as y_i is always positive,

$y_i \ln p_i$ will alway be negative.

So, the "-" sign makes the entropy a non-negative number. and appropriately measures the divergence between the true & predicted probability distributions, facilitating effective optimization during model training.

Q77. Explain overfitting and underfitting and how to combat them?

Q78. How do you combat curse of dimensionality.

Q79. Explain PCA.

Q80. Why ReLU better and more often used than Sigmoid in neural network?

Q81. Describe how Convolution works. What if your inputs are grayscale vs RGB imagery? What determines the shape of the next layer.

Q82. what is data normalization & why do we need it?

Q83. what is data normalization & why do we need it?

Preprocessing steps.

↳ rescale values to fit in specific range to assure better convergence during backprop.

if we don't scale features then high magnitude features will be weighted more in the cost function.

Q84. Why do we use convolutions for images rather than just FC layers?

Convolutions preserve, encode and actually use the spatial information from image.

FC layer would not.

Q85. What makes CNNs translation invariant?

each kernel acts as its own filter/feature detector.

so, while doing object detection it doesn't matter where your object is inside the image. Since we are using sliding window mechanism across the entire image.

Q86. Why do we have max pooling layer in CNNs?

helps to extract important features from input while simultaneously reducing dimensionality of the data.

Q87. What is batch normalization & why does it works.

Q88. Why would you use many small convolutional kernels such as 3×3 rather than a few large ones?

two reasons:

(1) Higher filter → more computational complexity & a higher no of parameters.

(2) Smaller kernel focuses more on each part of the input & improves its ability to recognize more complex structures while maintaining computational efficiency.

Q89. Why do we need a validation set and test set? What is the difference between them?

Q90. what is stratified cross-validation & when we should use it?

Q91. why do ensembles typically have higher scores than individual models?

Q92. what is an imbalance dataset? How to deal with it?

Q93. difference between supervised, unsupervised and reinforcement learning.

Q94. What is data augmentation? Can you give some examples?

technique to synthesize new data by modifying existing data.

Useful in Computer vision

Ex — Resize
Horizontal & vertical flip
Rotate
Add noise
Deform.

Q95. define precision, recall, F1-Score.

Q96. What is Cost function?

Q97. List different activation functions.

Q98. define learning rate.

Q99. Difference between Epoch vs Batch vs Iteration.

Epoch → one forward pass & one backward pass of all training examples.

Batch → Set of data processed in one pass (forw & back)

Iteration → no of training examples / Batch size.

Q100. What is vanishing gradient?

Q101 What are drop outs?

Q102 Define LSTM.

Q103 List the key components of LSTM.
Gates (forget, memory, update & Read)

tanh(x) (values between -1 to 1)
sigmoid(x) (values between 0 to 1).

Q104. List of variants of RNN.
LSTM, GRU,

Q105. What is Autoencoder? name a few applications?
It is basically compressed form of data.
few applications include —
→ Data denoising
→ Dimensionality reduction
→ Image reconstruction.

Q106. What are the components of GAN?
→ Generator
→ Discriminator

Q107. What's the difference between boosting & bagging?

Q108. Explain how ROC Curve works?

Q109. Explain the difference between type I & type II errors?

Q110. what is the difference between a generative and discriminative model?

Q111 When to use one-hot encoding & Label encoding?
One hot — no of categorical features are less & not ordinal data (like country)
Label Encoding — ordinal data, no of categorical features are large.

Q112 Difference between LDA & PCA for dimensionality reduction.

LDA → Supervised
PCA → Unsupervised.

Q113. What is t-SNE?

Q114. Basic different between LSTM & transformer.

Q115. What ~~are~~ the importance of Tokenization & challenges of it?

Importance — Text processing
Feature Extraction
Language Understanding.

Challenges — Ambiguity. (multiple meaning of same word).
out-of-vocabulary words.

Q116. How we can mitigate Bias in NLP?

It occurs when the data used to train or test an NLP model using bias dataset.

Why it is a problem?

↳ Let's say filtering resume based on gender, ethnicity etc.

Q117. Key evaluation matrics for assessing the performance of NLP models → for text classification and machine translation.

Accuracy / Precision / Recall / F1 score → classification

BLEU / ROUGE score → translation.

Q118. Explain the concept of word embeddings & how they are used in NLP models like word2vec & Glove.

Q119. Explain the concept of transfer learning in NLP and its importance in building effective NLP models.

Q120. discuss some common techniques to handle out-of-order vocabulary word (OOV) in NLP tasks.

(1) replace it with special token (<UNK>)
(2) character level Embeddings (Byte-Pair Encoding).

Q121. difference between supervised, unsupervised and semi supervised learning in NLP & when would you use each approach?

Supervised:
labeled data.

unsupervised:
topic modeling / word embeddings.

semi supervised:

Q122 In the context of NLP, what are some common techniques for text data preprocessing & cleaning?

(1) Lowercasing
(2) Tokenization
(3) stopwork removal
(4) Punctuation & special char removal.
(5) white space removal
(6) lemmatization
(7) stemming
(8) domain specific preprocessing.

Q123. Describe some popular libraries & frameworks used for working with NLP tasks, such as SpaCy, NLTK or HuggingFace Transformer.

SpaCy

POS tagging / NER / dependency parsing.

NLTK

Tokenization, Stemming, POS tagging, text classification.

HuggingFace Transformer

BERT / GPT.

Gensim

Word2vec, Glove, LDA.

Q124. Explain Seq-to-Seq models in NLP & their applications, especially in tasks like language translation & text summarization.

applications of seq-to-seq models:

→ language translation
→ Text summarization
→ Conversational agents & chatbots.
→ Q&A.

Q125 What are foundation models?

backbone of generative AI by learning versatile knowledge from massive dataset that allow them to support diverse capabilities via pretraining & customization.

Q126 Evaluating LLM & RAG systems - Metrics

① Ragas score :

generation

faithfulness → How factually accurate is the generated answer

answer relevancy → how relevant is the generated answer to the question

retrival

Context precision → the signal to noise ratio of retrieved context.

Context recall → Can it retrieve all the relevant information required to answer the question.

Metrices to evaluate the entire LLM or RAG system include:

→ Answer semantic similarity

→ Answer correctness.