Round 1:

The interview began with a formal introduction.:

1. PCA explanation with mathematical details.
2. Overfitting vs. underfitting(Scenario based question).
3. Vanishing gradient.
4. L1 and L2 regularization (with questions on why and how, focusing on mathematical intuition).
5. Neural network dropout.
6. Covariance (mathematical intuition).
7. Gradient boosting.
8. Comparison of Random Forest, Decision Tree, Gradient Boosting, and XGBoost (how they differ, with mathematical intuition).
9. How to handle data, outliers, standardization, and normalization.
10. Coding question: Find the number of possible palindromes from a string.
11. GRU.
12. Drawbacks of LSTM.
13. Different activation functions and how they behave differently.
14. Softmax.
15. Degree of freedom, entropy, and information gain.
16. Chi-square test.
17. Any questions for the panel members.

Round 2:

The questions included:

1. PCA in more depth.
2. Reconstruction loss.
3. Autoencoders (maths and logic explained).
4. LSTM in depth (issues with it and how to overcome them).
5. Can LSTM have a vanishing gradient issue?

6. What is the job of the forget gate?

7. Applying regression to a given dataset (mentioning all EDA steps).

8. Correlation.

9. Pearson's correlation coefficient.

10. How correlation and covariance differ (in-depth mathematical discussion).

11. Can transformers have a vanishing gradient issue?

12. What is the job of skip connections?

13. What is XGBoost? What is gradient boosting?

14. Explanation of the term "gradient" in gradient boosting (mathematical explanation).

15. The meaning of linear dependence and independence.

16. Why are decision trees prone to overfitting, and how would you solve that?

17. In a CNN, would you increase the depth of filters or neural layers? Why and how?

18. Discussion on height vs. depth in neural networks.

19. Non-linearity (which I could not answer).

20. How is context transferred in LSTM? How does the transformer solve that? (The interviewer provided some context.)

21. Calculate entropy (explain the formula).

22. How to compare different distributions.

23. Intuitive explanation of KL divergence.

Round 3:

1. Discussion on questions from other panel members.

2. What is correlation?

3. Pearson's correlation coefficient.

4. Other correlation metrics you know.

5. Rate yourself in Python, ML, and probability.

6. Discussion on BERT.

7. A probability question.

8. A puzzle (initially, I said I couldn’t solve it, but after being pushed to try, I was able to solve it in 5-6 minutes).

9. Given a dataset, how would you use it for a classification problem (from EDA to model selection)?

10. Different types of optimizers (how they work, with mathematical intuition).

11. What is an exploding gradient? Why does it happen? How do you solve the exploding gradient issue?

12. Some managerial questions related to estimation (e.g., What is the population of Bangalore? How many might have bikes and cars? How would that differ from your city? Do you know the child-to-adult ratio in your area?).

13. Discussion on the coding question from the first round (how I solved it, its time complexity, and whether it can be solved in a more optimized manner).

14. Further questions on correlation and covariance.

15. Different loss functions: MSE, MAE, cross-entropy.

16. When to use MAE, and why is it used?

17. What do you mean by differentiability?

18. Why did you rate probability lower than your ML skills? Are you underconfident?

19. Hypothetical questions: If I remove a specific component from the transformer, what will happen? If I remove a particular component from LSTM, what will happen?

20. Asked if I had any questions for the interviewer.

For any answer, if there was mathematical intuition available, the interviewer asked me to provide it. They also presented scenarios and asked how I would solve them using different methods, leading to a lot of cross-questioning for almost every answer. Unfortunately, not a single question was asked from my resume.