Interview Questions

→ Define LSTM

Ans LSTM is a type of RNN designed to handle the vanishing gradient problem, making it well-suited for tasks involving sequential data, such as time series analysis or NLP.
LSTMs can capture long-term dependencies in sequence, unlike traditional RNNs, which struggle to maintain information over longer time spans.

→ structure of LSTM

An LSTM consists of a series of gates that regulate the flow of information:

→ Forget Gate: Decides what information to discard from the cell state.
→ Input Gate: controls which new information is added to the cell state.
→ Cell State: carries the long-term memory, allowing information to persist over time.

→ output gate : Determines what the next hidden state will be, which is used for predictions.

Gates use sigmoid and tanh activation functions to control the flow of data.

Q. Error calculations metrics for LSTM

→ Mean Absolute Error (MAE)

↳ measures the average magnitude of errors in a set of predictions, without considering their direction

→ Mean Squared Error

↳ measures the average of squares of the errors which penalizes larger errors more than MAE

→ Root mean square error (RMSE)

↳ The square root of MSE giving the error in the original units of the target variable.

→ cross-Entropy error

↳ used in classification problems, measures the difference between the true label distribution and the predicted probability distribution.

Q. why LSTM over RNN?

vanishing Gradients: RNNs suffer from vanishing, making it difficult to learn long-term dependencies.

Q. how LSTM avoids overfitting

→ Regularization: L1, L2,

→ cell state mechanism

↳ memory cell allows LSTM to focus on more relevant information and ignore noise

→ Data Augmentation -

Q. can decision tree overfit?

Yes if:

→ tree is allowed to grow too deep without pruning

→ Tree captures noise or small fluctuations in the training data.

Q. what are Random Forests?

→ Ensemble learning method based on decision trees

→ A large number of DTs all trained on a different and random subset of data and features.

→ Final output is made by aggregating the predictions of all DTs

Q How does Random Forests avoid overfitting?

→ Averaging over many trees.
↳ reducing the variance of prediction

→ Bootstrap Aggregating.

↳ Each Tree is trained on a different bootstrap sample

→ Random Feature Selection

↳ At each node of tree a random subset of features is considered

Q. Difference between LSTM and GRU.

→ cell state → LSTMs have a seperate memory cell state and hidden state. whereas GRUs, combine these into one state.

→ Gates → LSTMs have 3 gates

↳ Input
↳ Forget
↳ Output

GRUs have 2 gates

↳ update
↳ Reset

→ GRUs are simpler and computationally more effective

9. R^2 and MR^2?

R^2 → metric used to assess how well the model's predictions match the actual data, provides the proportions of the variance in the dependent variable that is predictable from the independent variable

Use for small models

MR^2 → Adjusted R^2

↳ when multiple independent variables are involved and penalizes irrelevant data

For large models

9. Metrics used for classification

Accuracy: overall correctness of a model

Precision: Ratio of correctly predicted positive observations to total predicted.

Recall : Ratio of correctly predicted positive observations to the actual positives.

F1 score : harmonic mean of Precision and Recall.

Q. optimization in ML :-

→ Feature Engineering
→ Hyperparameter tuning
→ cross-validation

Q. Gradient Descent :-

optimization algo used to minimize loss function by iteratively adjusting model parameters weights in the the direction of the steepest decrease in the loss function to find the optimal solution