1. What is overfitting and how to avoid it?

1. Overfitting refers to a model that closely aligns to a particular data set of points. We can avoid it by regularization, cross-validation, ensembling methods, early stopping etc.,

2. What is RMSE and MSE? How can you calculate them?

1. Mean Squared Error (MSE) and Root mean squared error (RMSE) are two of the most common metrics used to measure accuracy for continuous variables. They show deviation of the residuals or how far the points are from the regression or modelled line. The mean of the sum of squares of the difference between actual and predicted gives mse and the root of mse gives rmse.

3. What is Line of best fit?

1. The line that best fits the given set of points where the loss is least.

4. Explain multivariate linear regression using a real-life example.

1. prices of grains in agricultural markets vary every day. You may imagine that it's daily price(Yt) fluctuations depend on last day's temperature(Tt-1), last day's humidity(Ht-1), last day's sold out stock(St-1), last day's market arrivals(At-1), last day's price of substitute commodity(Ct-1) etc. You can make following multiple regression equation---

Yt = w0 + w1\*Tt-1 + w2\*Ht-1 + w3\*St-1 + w4\*At-1 + w5\*Ct-1 + error

5. How can we improve the accuracy of a linear regression model?

1. Add more data, remove outliers, ensemble methods, algorithm tuning, feature selection.