

Minor project 1

1. What is overfitting and how to avoid it?

Overfitting is when a statistical model fits exactly against its training data. When this happens, the algorithm unfortunately cannot perform accurately against unseen data, defeating its purpose.

Ways to avoid it:

1. Early stopping: This method seeks to pause training before the model starts learning the noise within the model.
2. Train with more data
3. Data augmentation

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

Mean Squared Error represents the average of the squared difference between the original and predicted values in the data set. It measures the variance of the residuals.

$$MSE = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2$$

Root Mean Squared Error is the square root of Mean Squared error. It measures the standard deviation of residuals.

$$RMSE = \sqrt{MSE} = \sqrt{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y})^2}$$

3. What is Line of best fit?

A **Line of best fit** is a straight line that represents the best approximation of a scatter plot of data points. It has the minimum MSE or RMSE.

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

Multivariate linear regressions involve **multiple data variables** for analysis. In the real world, there are many situations where many independent variables are **influenced by other variables** for that we have to move to different options than a single regression model that can only take one independent variable.

Multivariate Linear Regression helps us to measure these angles of more than one independent variable and more than one dependent variable. It finds the **linear relation between the variables**.

For example Housing Prices which will depend on area of house, locality, if it is furnished or not etc.

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

Ways to improve accuracy of a linear regression model are:

1. Normalising data, i.e., shifting it to have a mean of zero, and a spread of 1 standard deviation
2. Regularisation of the features to reduce possible overfitting