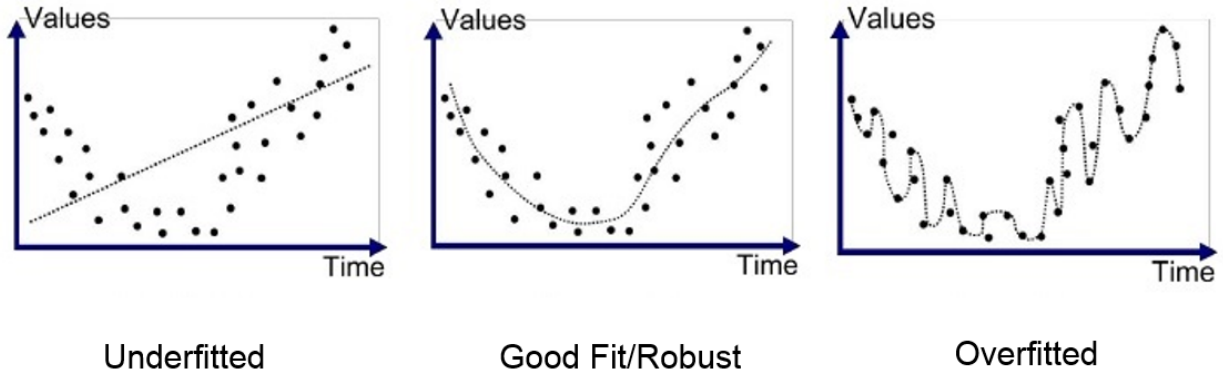
Name : Priya Mohata

College : PES University

MINOR-PROJECT-1 Theory Questions

1. What is overfitting and how to avoid it?

- If our model does much better on the training set than on the test set, then we're likely overfitting.
- Overfitting happens when a model learns the details in the training data to the extent that it negatively impacts the performance of the model on new/test data.
- For example, it would be a big red flag if our model saw 99% accuracy on the training set but only 55% accuracy on the test set.
- Prevention of overfitting :
    - Cross Validation
    - Adding more data
    - Remove features



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

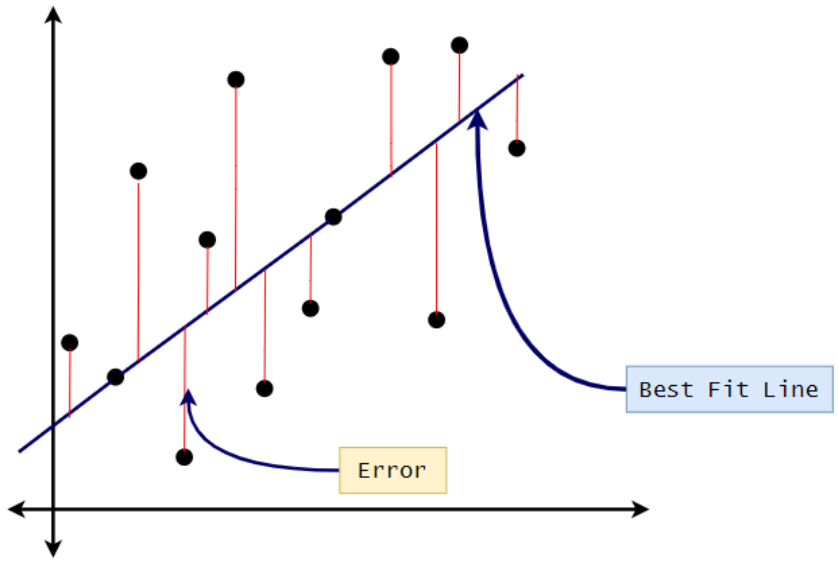
- MSE : Mean Squared Error represents the average of the squared difference between the original and predicted values in the data set. It measures the variance
- RMSE is the square root of Mean Squared error. It measures the standard deviation

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

$$RMSE = \sqrt{MSE}$$

3. What is Line of best fit?

Line of best fit refers to a line through a scatter plot of data points that best expresses the relationship between those points



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

- Multivariate Regression is a type of machine learning algorithm that involves multiple data variables for analysis.

- If Amazon has collected the data of customers like age , purchase history, gender,amazon pay balance , it might want to find relationships between these dependent and independent variables.

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

1. Add more data
2. Treat missing and Outlier values
3. Feature Selection
4. Algorithm Tuning
5. Data Visualization