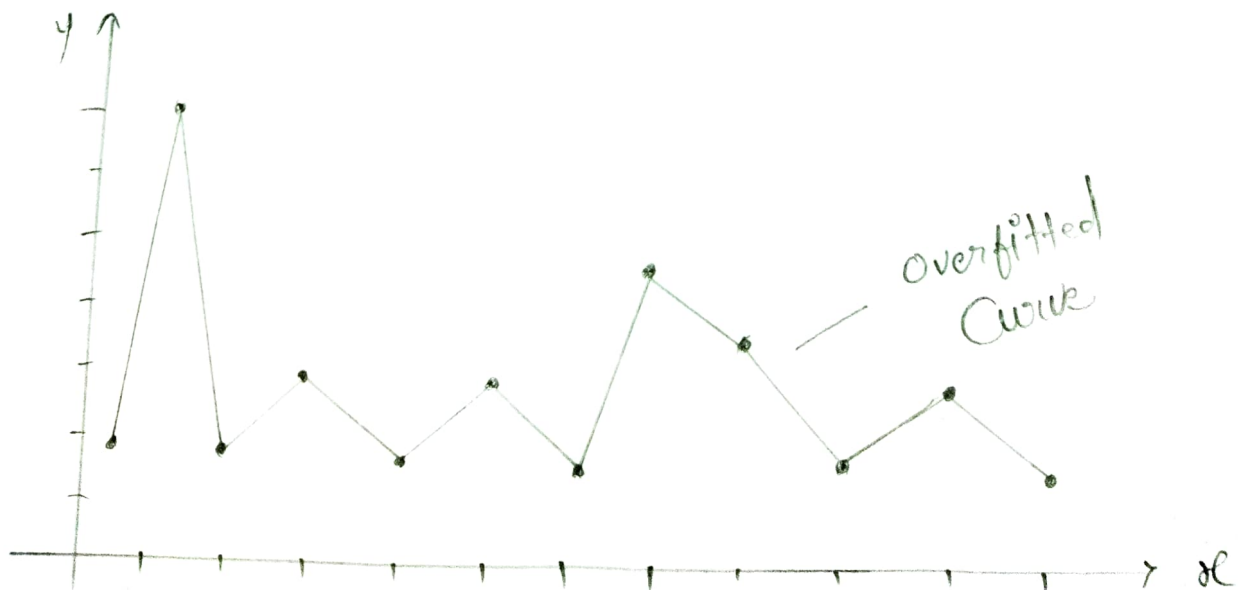


Name : Mangesh Khode
Class : B.Tech 3rd sem

What is Overfitting?

When a model performs very well for training data but has poor performance with test data; it is known as overfitting. In this case, the machine learning model learns the details and noise in the training data such that it negatively affects the performance of the model on test data. Overfitting can happen due to low bias and high variance.

Example : The concept of the overfitting can be understood by the below graph of the linear regression output:



As we can see from the above graph, the model tries to cover all the data points present in the scatter plot. It may look efficient, but in reality, it is not so. Because the goal of the regression model is to find the best fit line, but here we have not got any best fit so, it will generate the prediction errors.

Reasons for Overfitting :

- * Data used for training is not cleaned and contains noise (garbage values) in it.
- * The model has a high variance.
- * The size of the training dataset used is not enough.
- * The model is too complex.

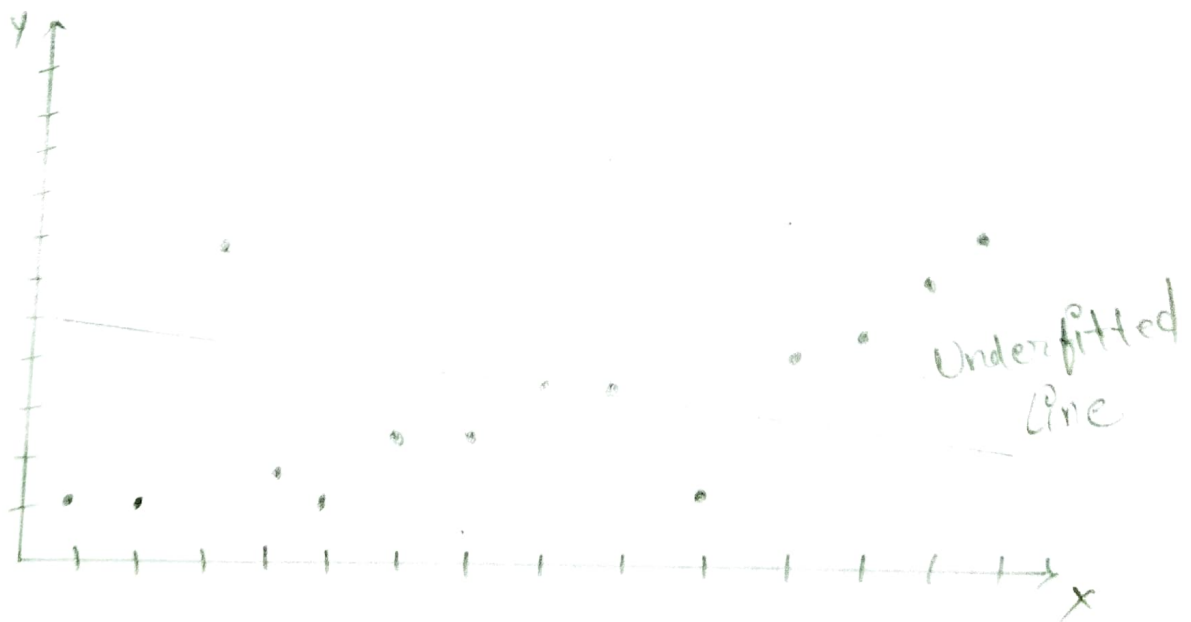
Ways to Tackle Overfitting :-

- * Using K-fold cross-validation
- * Using Regularization techniques such as Lasso and Ridge.
- * Training model with sufficient data
- * Adopting ensemble techniques.

What is Underfitting?

When a model has not learned the patterns in the training data well and is unable to generalize well on the new data, it is known as underfitting. An underfit model has poor performance on the training data and will result in unreliable predictions. Underfitting occurs due to high bias and low variance.

Example :- We can understand the underfitting using below output of the linear regression model:



As we can see from the above diagram, the model is unable to capture the data points present in the plot.

Reasons for Underfitting

- * Data used for training is not cleaned and contains noise (garbage values) in it.
- * The model has a high bias
- * The size of the training dataset used is not enough
- * The model is too simple.

Ways to Tackle Underfitting :

- * Increase the number of features in the dataset
- * Increase model complexity
- * Reduce noise in the data.
- * Increase the duration of training the data.