Over vs Under vs Best Fitting

Overfitting: Good performance on the training data, poor performance on test data.
Underfitting: Poor performance on the training data and poor performance on test data.

Must Read this blog!

Dimensionality Reduction

In simple words, dimensionality reduction refers to the technique of reducing the dimension of a data feature set.

Why Dimensionality Reduction?

dimensionality reduction has many other benefits, such as:

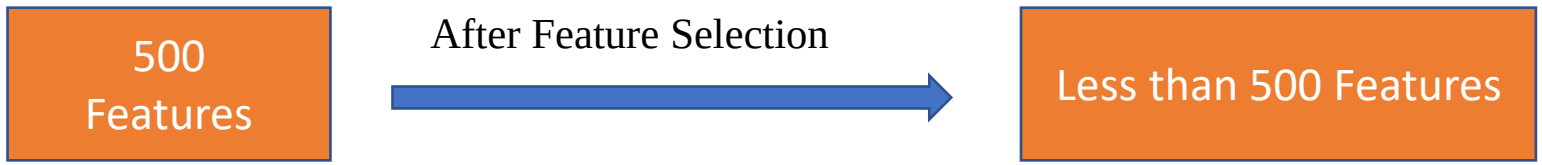
- It eliminates noise and redundant features.
- It prevents overfitting and underfitting in ML models.
- It helps improve the model's accuracy and performance.
- It facilitates the usage of algorithms that are unfit for more substantial dimensions.
- It reduces the amount of storage space required (less data needs lesser storage space).
- It compresses the data, which reduces the computation time and facilitates faster training of the data.

Dimensionality Reduction Techniques in ML

- Feature selection.
- Feature extraction.
- Principal Component Analysis (PCA)
- Non-negative matrix factorization (NMF)
- Linear discriminant analysis (LDA)
- Generalized discriminant analysis (GDA)
- Missing Values Ratio.
- Low Variance Filter.

Feature Selection

Feature Selection is the process where you automatically or manually select those features which contribute most to your prediction variable or output in which you are interested in. Having irrelevant features in your data can decrease the accuracy of the models and make your model learn based on irrelevant features.



All about Feature Selection

Feature Selection is the process where you automatically or manually select those features which contribute most to your prediction variable or output in which you are interested in. Having irrelevant features in your data can decrease the accuracy of the models and make your model learn based on irrelevant features.

All about Feature Selection

Techniques:-

- Feature Importance
- SelectKBest

All about Feature Selection

Feature Importance: ExtraTreesClassifier()

Extra Trees Classifier: Extra Trees Classifier is an ensemble learning method fundamentally based on decision trees. ExtraTreesClassifier, like RandomForest, randomizes certain decisions and subsets of data to minimize over-learning from the data and overfitting

All about Feature Selection

SelectKBest: Score Function

- f_regression: Used only for numeric targets and based on linear regression performance.
- f_classif: Used only for categorical targets and based on the Analysis of Variance (ANOVA) statistical test.