**Cross validation-**

In cross validation, we use a subset of the datasets to test and train a model on different iterations. It is a resampling method in machine learning.

We often use cross validation when goal of the model is to make prediction and we want to know how accurately the model is working. We select the subset of the data randomly.

**Types of cross validation-**

1. K- fold cross validation
2. Monte- Carlo
3. Holdout cross validation

In k- fold cross validation, you take the data and partition it into k parts, k can be any integer value at a time. K=1 so one test data and the left over is for training. You train the model and get the result and repeat the same process k number of times.

Oversampling or upsampling- It involves selecting examples from the minority class with replacement and adding them to the training dataset.

Downsampling or undersampling- It is a method used to decrease the size of a dataset by removing some of the instances. This is often done to reduce computational complexity and training time and eliminate redundancy or irrelevant instances from the dataset.

Overfitting- Overfitting is a phenomenon in machine learning where a model learns the specific patterns and noise in the training data to such an extent that it negatively impacts its ability to generalize to unseen, new data. In essence, the model becomes overly complex and fits the noise in the training data rather than capturing the true underlying patterns and relationships.

Here are the key characteristics and implications of overfitting:

1. Overly Complex Model:

- An overfit model is often overly complex, capturing not only the true relationships in the data but also noise and outliers specific to the training set.

2. Poor Generalization:

- The overfit model performs very well on the training data (high training accuracy) but poorly on unseen data (low validation or test accuracy).

3. Memorization vs. Generalization:

- The model effectively memorizes the training data instead of learning the underlying generalizable patterns. It fails to generalize to new, unseen data.

4. High Variance:

- Overfitting leads to high variance in the model's predictions. Small changes in the training data can result in significantly different models and predictions.

5. Detecting Overfitting:

- One way to detect overfitting is by comparing the model's performance on a validation set (or a separate test set) to its performance on the training set. If the performance gap is significant, it's likely overfitting is occurring.

6. Preventing Overfitting:

- Regularization techniques like L1 and L2 regularization, early stopping, cross-validation, reducing model complexity, and using more data can help prevent or mitigate overfitting.

7. Bias-Variance Tradeoff:

- Overfitting is part of the bias-variance tradeoff. Models with high complexity tend to have low bias but high variance, making them susceptible to overfitting.

To combat overfitting, it's essential to strike a balance between model complexity and model generalization. Regularization and validation techniques are commonly used to ensure that the model doesn't overfit and can effectively make accurate predictions on new, unseen data.

Underfitting- Underfitting is a concept in machine learning where a model is too simple to capture the underlying patterns and relationships present in the data. In essence, the model is unable to learn the complexities of the data and performs poorly on both the training data and unseen, new data.

Here are the key characteristics and implications of underfitting:

1. Too Simple Model:

- An underfit model is often overly simplified and cannot capture the true relationships and patterns in the data.

2. High Bias:

- The underfit model has high bias, as it makes strong assumptions or simplifications about the data, which may not reflect the true data distribution.

3. Poor Performance:

- The model performs poorly on both the training data (low training accuracy) and unseen data (low validation or test accuracy). It fails to capture the underlying structure of the data.

4. Oversimplification:

- Instead of capturing the actual trends and patterns in the data, the model might oversimplify the problem, resulting in inaccurate and inadequate predictions.

5. Overgeneralization:

- Underfitting can lead to overgeneralization, where the model assumes that all data points are similar and applies a simple pattern across the board.

6. Detecting Underfitting:

- Low training and validation/test accuracy, along with a lack of improvement in performance even with more training, are indicators of underfitting.

7. Addressing Underfitting:

- To address underfitting, you can try using more complex models, incorporating more features, increasing the model's capacity, or using advanced algorithms that can capture the complexities of the data.

8. Bias-Variance Tradeoff:

- Underfitting is part of the bias-variance tradeoff. Models with high bias tend to have low variance, but they may not capture the true underlying patterns in the data.

Balancing model complexity is essential to achieve good performance. Underfitting represents an extreme on the side of overly simplistic models. Techniques such as increasing model complexity, fine-tuning hyperparameters, and providing more relevant features can help alleviate underfitting and enable the model to better learn the true relationships in the data.