SDAI ICE\_7

Report

Name : Kothapalli Phani Sagar

ID : 11555373

Github link : https://github.com/reshmi611/ICE\_7.git

1.According to you, why do overfitting and underfitting occur, and how resolve them? What is the difference between them?

A : When a machine learning model is unable to learn by capturing the underlying data pattern, underfitting occurs. A learning model that is underfitted will have a high bias, a low variance, and perform badly on unobserved data.

Contrarily, overfitting occurs when the learning model commits to memory the correlation between the input and target variables. A poorly performing overfit model will have low bias and large variance.

To create a generalized model with appropriate performance, you must balance underfitting and overfitting (a trade-off between bias and variance).

Basically, if you make your model more complex, you risk having an overfit model since you lose bias but gain variance. An underfit model runs the risk of having low variance and high bias. When a model is unable to capture the underlying pattern of the data, underfitting occurs. These models typically have low variance and strong bias. It occurs when there is insufficient data to develop a reliable model or when we attempt to fit a linear model to nonlinear data. Additionally, these types of models, like logistic and linear regression, are quite straightforward to use in capturing the complicated patterns in data.

### 2. What kind of pattern did you analyze in the Train and Test score while running the code of overfitting?

A: Overfitting occurs when a machine learning model becomes so adept at remembering the pattern in training data that it is unable to categorize unobserved data. The model learns from the noise or fluctuations in the training data, which it interprets as features. As a result, the model performs better on the training set than it does on the validation and testing sets.

The model learns from the noise or fluctuations in the training data, which it interprets as features. As a result, the model performs better on the training set than it does on the validation and testing sets.

If this pattern is seen, training for incrementally learning algorithms should end when performance on the training set degrades.

### 3. What is cross-validation, and what did you analyze in a different type of validation that you performed?

A: A statistical technique called cross-validation is used to gauge the effectiveness (or accuracy) of machine learning models. It serves as a safeguard against overfitting in predictive models, especially when the available data may be scarce. In cross-validation, you divide the data into a predetermined number of folds (or partitions), analyze each fold individually, and then average the total error estimate.

There are different types of cross validation methods, and they could be classified into two broad categories – Non-exhaustive and Exhaustive Methods. We’re going to look at a few examples from both the categories.

1) K-fold - One strategy to enhance the holdout method is K-fold cross validation. This approach ensures that our model's score is independent of how we chose the train and test sets.

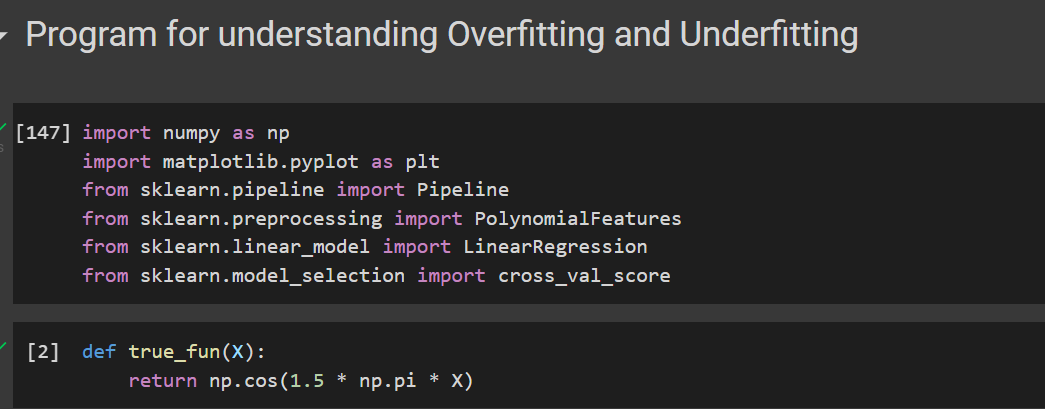
2) Repeated K-Fold

3) Leave One Out (LOO)

### 4. Explain the analysis from generated ROC and validation curve and what they represent?

A : The relationship or trade-off between clinical sensitivity and specificity for each feasible cut-off for a test or set of tests is usually depicted graphically using ROC curves.

You must understand the ideas of true positive, true negative, false positive, and false negative in order to create a ROC curve. When contrasting test results with clinical reality, these ideas are applied. The diagonal line is where the ROC curve of the useless test lies. It contains the point with a sensitivity and specificity of 50% each. The useless test's ROC curve's area under the curve is 0.5.



Graphical user interface, chart, histogram

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated