Task-2
Report for SDAI ICE-7 (Under-fitting and Over-fitting)

The goal of a good machine learning model is to generalize well from the training data to any data from the problem domain. The reason that over-fitting and under-fitting occurs when there is less data or heavy data that we fits to the model. When we fit the data to our machine learning model and that data is to less then over-fitting will occurs and when there is more excess data which can not fits to our model exactly then there are chances of under-fitting.

To resolve the over-fitting and under-fitting we can use more complex model also we can increase the features and reduce the noise and also use more training data so that we can overcome the over & under fitting models.

The incorrect predictions that initially make by our model accurately is also called as “Under-Fitting” and when that model does not make the accurate predictions to the data then it is called as “Over-Fitting”.

While running the code of over-fitting I analyzed that the trained model gave us the best result for the data and while testing it we can see the poor result than our train model this kind of scenario can be said as over-fitted as the predicted results are so poor than the trained result. We can say that the model is perfect when there will be good results in both the train and predicted data then that model is perfect.

Cross Validation will be used for evaluating the machine learning models by training the several machine learning models on it’s subset by the available input data to know the complementary subset of the data we are evaluating them.

By using the Cross validation we can detect over-fitting by performing the different validation types we can observe that there will be difference in the output accuracy where each validation iterator sets differently to the data so there will be changes in it’s training and output predicting accuracy.

Where the validation curve mainly represents how much our model goes to over-fitting and how much it goes to under-fitting with both the train and predicted values where the validation curve is a graphical technique that can be used for single hyper-parameter to measure it’s influence.

The ROC curve will show the graphical performance of a classification model at all the classification thresholds where the curve plots the True positive rate and False positive rate with the data. Where in the ROC Curve the curves that are closer to the top-left corner indicate a better performance.