Mohd Talha Patrawala
CMPN-B
23102B0025

ML: Assignment - 2

Q.1.

① Training data error only shows memorization not learning, while evaluation on unseen data estimates the model's generalization ability and prevents selection of models that only performs well on previously seen data.

② k-fold cross validation ensures that every data point is used both for training and validation, reducing the dependence on a single random split in hold-out validation. As a result, k-fold cross validation provides a more reliable and less biased estimate of the model's generalization performance.

Q.2.

① Model A is likely to overfit as it has higher variance than model B as it suffers from higher generalization error and break more easily when data changes.

② Generalization error is the difference between training performance and performance on unseen data.

③ Overfitting can be reduced by using regularization or a simpler model. Another effective approach is to increase the training data or removing irrelevant features.

Q.3.

① The model learns patterns from the old distribution, so when the input distribution changes, those patterns may no longer be valid which will reduce generalization performance.

② Complex models have low bias but high variance, making them sensitive to distribution changes and less stable. Whereas, Simple models have slightly higher bias but lower variance, so they are more stable when the input distribution changes.

③ Use cross validation periodically to check if performance is stable across different data splits and control model complexity using regularization or a simpler model to improve robustness.

Q.4.

① The model overfits for students with strong math as the difference between training error and test error is high. It underfits for students with weak math since the difference is low.

② This can be improved by using more balanced subgroup data and applying better group-aware features.

③ Use stratified train-test splits that preserve subgroup proportions and subgroup-specific performance metrics instead of relying on a single overall accuracy, which can hide poor performance for specific groups.