ML ASSIGNMENT 6

Q1) model is a file that can recognise data patterns, can perform machine learning tasks like classification and regression and has been trained to do those tasks. Best way to train a model is to understand use case and hardware properly, create proper dataset and select right hyper parameters for training.

Q2) no free lunch means that there is no best machine learning algorithm or optimiser and hence proper hyper parameters have to be selected based on use cases.

Q3) machine learning models can be evaluated and improved by splitting dataset into k batches, performing training on them separately and comparing with other batches. This can be used especially when there is limited data.

Q4) bootstrap sampling method involves creating samples with replacement from a dataset to understand about a population.

Q5) kappa value measures level of agreement between two classifications.

Q6) in ensemble method, there are multiple decision makers. In machine learning, results from several models are used to give best possible output.

Q7) descriptive model tries to describe an object or system in its entirety including all its relationships. It can be used in computer vision (relation between person and vehicle for helmet detection) and nlp (in completely understanding a word, its name entity and grammar).

Q8) r score, r square score etc.

Q9) descriptive model provides an outcome using data directly while predictive model predicts based on its understanding of patterns in data and its training. Underfitting occurs when model is unable to capture relations in data with decent accuracy while overfitting occurs when model is performs very poorly in testing when compared to training showing its lack of regularity and overly relying on training trends. Bootstrap sampling method involves creating samples with replacement from a dataset to understand about a population while machine learning models can be evaluated and improved by splitting dataset into k batches, performing training on them separately and comparing with other batches with cross validation.

Q10) in LOOCV, one splitted data is considered as validation set and rest of splitted data is treated as training set. F score measures accuracy of binary classification. Silhouette index also measures accuracy of classification and also clusters. ROC curve shows performance of classification model at various thresholds.