

# CatBoost

## THEORY

- 1) CatBoost or Categorical Boosting is an open-source boosting library developed by Yandex. In addition to regression and classification, CatBoost can be used in ranking, recommendation systems, forecasting and even personal assistants.
- 2) The greedy approach takes an average of the target for a category group. But it suffers from target leakage as the target value is being used to calculate a representation for the categorical variables and then using those features for prediction.
- 3) The Holdout method tries to reduce this by partitioning the training data set. But this significantly reduces the effective use of training data.
- 4) This technique also ensures the use all the available past for each example to compute its target statistics and thereby encoding the categorical variables.

## QUIZ

- 1) What are the advantages of CatBoost?
- 2) What are the disadvantages of CatBoost?

## ANSWER

- 1) It is easy to use with packages in R and Python, also handles categorical features effectively by ordered target statistics.
- 2) It can perform very bad if the variables are not properly tuned & performs better than other algorithms only when we have categorical data.