**Assumption of CatBoost Algorithm**

<https://catboost.ai/en/docs/>

|  |  |  |
| --- | --- | --- |
| S No | Assumptions | Description |
| 01 | Train model on CPU & GPU Processing | Training on GPU requires NVIDIA Driver of version 450.xx or higher. |
| 02 | Use for Numerical & Categorical Data | Import libraries  For Numerical Data = from catboost import CatBoostRegressor  For Categorical Data = from catboost import CatBoostClassifier |
| 03 | Regular Prediction | **CatBoostClassifier** – method predict\_proba  Apply the model to the given dataset to predict the probability that the object belongs to the given classes  **CatBoostRegressor** – method predict  Apply the model to the given dataset. |
| 04 | Cross-Validation (cv) | Training can be launched in cross-validation mode. In this case, only the training dataset is required. This dataset is split, and the resulting folds are used as the learning and evaluation datasets. |
| 05 | Usage | Can be used in Python & R Packages |
| 06 | Using Overfitting Detector | If overfitting occurs, CatBoost can stop the training earlier than the training parameters dictate. For example, it can be stopped before the specified number of trees are built. This option is set in the starting parameters. |
| 07 | Pre-trained Data | Initial formula values can be set for all input objects in the Python package. The training starts from these values for all input objects instead of starting from zero. |
| 08 | Calculate Metrics | A list of specified metrics can be calculated for the given dataset using the Python package (CatBoostRegressor, CatBoostClassifier) – method use eval\_metrics |
| 09 | Categorical Features | **Warning** Do not use one-hot encoding during preprocessing. This affects both the training speed and the resulting quality. Because, by default, CatBoost uses one-hot encoding for categorical features with a small amount of different values in most modes.  CatBoost supports numerical, categorical, text, and embeddings features. Categorical features are used to build new numeric features based on categorical features and their combinations.  **Parameter**  **Command-line version parameters:** --one-hot-max-size  **Python parameters:** one\_hot\_max\_size  **R parameters:** one\_hot\_max\_size |
| 10 | Text Features | CatBoost supports numerical, categorical, text, and embeddings features. Text features are used to build new numeric features. |
|  |  |  |