|  |  |
| --- | --- |
| 2021/02/28 | Validation |
| **k-Fold Cross-Validation** | Cross-validation is a statistical method used to estimate the skill of machine learning models.  The procedure has a single parameter called k that refers to the number of groups that a given data sample is to be split into.  The general procedure is as follows:   * Shuffle the dataset randomly. * Split the dataset into k groups * For each unique group:   + Take the group as test data set   + Take the remaining groups as a training data set   + Fit a model on the training set and evaluate it on the test set   + Retain the evaluation score and discard the model   Example: k = 5  The train dataset will be divided into 5 folds   |  |  |  |  | | --- | --- | --- | --- | | Epoch | Train data | Test data | Result (accuracy) | | 1 | Fold 2 - 3 - 4 - 5 | Fold 1 | 90% | | 2 | Fold 1 - 3 - 4 - 5 | Fold 2 | 80% | | … |  |  |  | | n = k = 5 | Fold 1 - 2 - 3 - 4 | Fold 5 | 85% | |
| Quote |  |