

An Analysis of Model Evaluation with Cross-Validation: Techniques, Applications, and Recent Advances

Jinhui Qiu^{1,a,*}

¹*Mathematics Department, Santa Monica College, Santa Monica, CA 90405, The United States*

a. qiu_jinhui@student.smc.edu

**corresponding author*

Abstract: Cross-validation is the reuse of data, and then the acquired sample data is divided into different training sets and test sets. The model is trained with the training set, and the model forecast quality is evaluated by the test set. Based on these results, this paper can get several different training sets and test sets. A sample in a training set may become a sample in the next test set. This paper focuses on the three key cross-validation approaches. Recent advances such as nested cross-validation for model selection and time series cross-validation for sequence data are also discussed. For instance, in the area of home price forecasting, k-fold cross-validation ensures that the model performs reliably over different data segments, thus proving its robustness. Also, the geographical information data set is an example, in which the location of the data points is closer, the more dependent they are. By synthesizing insights from various studies, this review provides a comprehensive understanding of how cross-validation techniques can enhance model evaluation and guide the development of more accurate prediction models.

Keywords: Cross-Validation, Model Evaluation, Machine Learning.

1. Introduction

Cross-validation is one of the fundamental techniques in machine learning, and it is widely applied to estimate the prediction error [1]. The key idea of this approach is to divide the data set into several subsets or "folds", then train the model on certain folds and test them on the rest of the folds. This system makes sure that every data point has a chance to be present in both the training set and the test set, thereby providing a robust evaluation of the performance of the model.

Cross-validation is mainly aimed at avoiding overfitting, that is, the model is good for training data, but not for new data. A common cross-validation method is k-fold cross-validation. K-fold cross-validation is often proposed to overcome the difficulty of obtaining a proper classification and a good performance estimation. It splits the data into equal sizes of k components. The model was trained k times, and the remaining 1-fold was tested [2]. For example, in 5-fold cross-validation, the dataset is divided into 5 parts. The model is trained in 4 sections, and the rest is tested in one section. Then, all 5 sections are rotated in order to form a test group. This technique ensures that the performance of the model is evaluated on different data splits, making it one of the preferred methods for validating machine learning models. By using cross-validation, it can gain insight into how the model performs on unseen data, thereby achieving more accurate and reliable predictions.

2. Literature Review

Cross-validation is of great importance to build a machine learning model by comparing and selecting suitable models and evaluating their forecast performance. This process is not restricted to the training stage of the model but also includes the evaluation of the model's performance on the new dataset (test data). Since its birth, Cross-Validation has developed significantly and become the cornerstone of the evaluation model in the field of Machine Learning. It originated in the early days of statistical learning when the need to estimate the performance of a model on unseen data was first recognized. Several cross-validation techniques have been developed to improve the reliability of these estimates over the years.

K-fold cross-validation is easy to construct, adaptable, and less dependent on the structure of the model. The K-fold Cross-Validation method divides the dataset into k-comparable subsets or folds. The partition folding was trained and tested in K iterations, and the rest of the K-1 folding model was tested and trained with a folding method. The most common K values are 5 and 10 [3].

Leaf One Out Cross Validation (LOOCV) is widely used in structural optimization and reliability analysis. LOOCV is just one sample for each subset. Compared with other CV methods, LOOCV has been widely used in agent-based optimization due to its ability to provide approximate unbiased error estimates [4]. Although LOOCV can take up a lot of time, it is especially useful for small datasets, since it maximizes the size of a Training Set per iteration [5].

Compared to the conventional Holdout Validation, which is a one-time-split between Training Set and Test Set, cross-validation provides a more complete assessment by evaluating the model on multiple sections of data. This reduces the risk of bias assessment that may occur in a single Train-Test Split. The Bootstrap method provides another alternative through resampling with replacement but may introduce higher performance estimation variance than Cross-Validation.

Although Cross-Validation has many advantages, it is not without challenges. Historically, Cross-Validation has played an important role in validating models in various fields. The research highlights how k-fold Cross-Validation can help select the best model by providing more reliable performance metrics than a simple Train-Test Split. In modern applications, Cross-Validation is still critical in tuning hyperparameters and selecting models. For example, in geographic information system (GIS) data sets, k-fold cross-validation (CV) can be used to verify that the assumed data is independent. The closer the data points are geographically, the stronger the dependence between them is. This phenomenon, known as Spatial Auto-Correlation (SAC), can result in the estimation of the performance of a spatial model by standard cross-validation (CV). To overcome this problem, SKCV provides a useful estimation of the performance of the model prediction without the positive bias due to SAC [6].

3. Methodology

In Machine Learning, cross-validation has been widely applied to evaluate model performance. The following are several standard and advanced cross-validation techniques, as well as their applications and characteristics.

K-fold cross-validation is one of the most commonly used methods for cross-validation. In this approach, the data set is divided into k-folds of equal size [7]. Each time, k-1 replicates are selected as the training set, and the other 1 replicate are used as test sets. After this round is completed, randomly re-selected copies to train the data. After a few cycles ($< k$), the optimum model and parameters are evaluated by selecting the loss function. The reliability and veracity of the results of cross-validation are largely dependent on the value of k. Some scholars have used the credit card customer dataset in the machine learning library of the University of California, Irvine to study the delinquency of credit card customers in Taiwan. The data contains 30,000 observations. The results

show that in most cases, the K-Fold Cross-Validation method achieves a lower MSPE average than other model selection and model averaging methods, achieving lower risk [8].

Stratified K-Fold Cross-Validation is better than K-Fold. The comparison is done by maintaining the proportion of samples in each category [9]. This approach can be useful for unbalanced datasets (such as class imbalance in classification tasks), as it ensures that the distribution of data in every fold is similar to that of the whole dataset. This is helpful for the generalization of the model, especially in classification tasks. This CV object returns a hierarchical collapse and is a variant of K-Fold.

Leave-One-Out Cross-Validation requires the ability to break down the observed model into simple terms [10]. In LOOCV, each data point is used as a test set in a separate training test cycle, while the remaining data points are used as training sets. This method maximizes the size of the training set in each iteration and is suitable for small data sets. However, since only one data point is retained for testing in each iteration, the computational cost is high and it is easy to produce high variance. Christel et al. used data from ovarian cancer patients and focused their research on overall survival [11]. They used LOOCV to predict the survival time of high-risk and low-risk patients. The results showed that the median survival of the low-risk group was 11.6 months. The median survival of the high-risk group was 4.2 months. From the research method point of view, for smaller samples, the effect of validating Cox PH models using a single split sample or external data sets is not as good as LOOCV.

4. Results

In the model performance evaluation of machine learning, different Cross-Validation methods have a significant impact on the accuracy and reliability of the results, as shown in Table 1.

Table 1: Cross-Validation Method

Technology	Operation steps	Advantages
K-Fold Cross-Validation	The entire dataset is divided into K equal-sized parts.	The entire dataset is used as both a training set and a validation set.
Stratified K-fold cross-validation	Similar to K-fold cross-validation, but in each fold the class proportions are the same as in the entire dataset.	It works better when dealing with imbalanced datasets, ensuring that the distribution of classes in each fold is similar.
Leave-One-Out Cross-Validation	One sample point is used as a validation set, and the remaining n-1 samples are used as a training set.	All data samples are used as training and validation samples.

For example, K-Fold Cross-Validation provides stable performance estimation because it reduces the randomness of the results by training and testing the model multiple times. This method has been proved in many studies to effectively balance the evaluation of model performance so that it has both computational efficiency and robust results. There are many ways to divide the data. For example, stratified sampling can be used for unbalanced data, which means that the same category ratio as the original data set is maintained in each subset.

Cross-Validation also plays a key role in handling Bias-Variance Tradeoffs. Less Folds (such as 3 fold) may lead to high variance, while more Folds (such as 10 fold) may better capture the overall performance of the model performance, but the computational cost will also increase. Leave-One-Out Cross-Validation (LOOCV) has the advantage of maximizing the utilization of training data, but it can result in high variance, particularly for small datasets.

Cross-Validation technology has shown its importance in practical applications in various fields. For example, in the financial field, Cross-Validation is used to evaluate and optimize trading algorithms to ensure their robustness under different market conditions. In the field of health care, Cross-Validation is widely used in the evaluation of disease prediction models to ensure that the model has good generalization ability in different patient groups.

5. Conclusion

Generally speaking, cross-validation is one of the key techniques in Machine Learning. Several commonly used cross-validation methods are reviewed, such as K-Fold Cross-Validation, K-stratification, and Leave-One-Out Cross-Validation (LOOCV). Through these methods, Overfitting can be effectively avoided and reliable estimation of model performance can be obtained. Cross-validation can verify the model multiple times, reducing the deviation in the model error and making the accuracy of the validation set more reliable. Cross-validation can also obtain multiple models, and multiple predictions can be made on the test set to increase the diversity of the prediction results.

Although Cross-Validation provides many advantages, such as reducing the bias of model evaluation and improving the robustness of results, it also has some shortcomings. For example, although LOOCV can maximize the use of training data, it is too expensive to calculate on large data sets and easily leads to high variance. Although K-Fold Cross-Validation and Stratified K-Fold balance the computational efficiency and reliability of the results, they still have limitations in processing time series data. It needs to consider comprehensively and select the optimal solution.

References

- [1] W. Jiang, & R. Simon, *A comparison of bootstrap methods and an adjusted bootstrap approach for estimating the prediction error in microarray classification*. *Stat in Med*, 29 (2007), 5320-5334.
- [2] O. Oyedele, *Determining the optimal number of folds to use in a K-fold cross-validation: A neural network classification experiment*. *Math. Res.*, 1 (2023).
- [3] A. M. Peco Chacón, I. Segovia Ramírez, & F. P. García Márquez, *K-nearest neighbour and K-fold cross-validation used in wind turbines for false alarm detection*. *SFC* (2023).
- [4] Y. Pang, Y. Wang, X. Lai, S. Zhang, P. Liang, & X. Song, *Enhanced Kriging leave-one-out cross-validation in improving model estimation and optimization*. *Comput. Methods Appl. Mech. Eng.* (2023).
- [5] D. Du, K. Li, M. Fei, & G. W. Irwin, *Automatic forward model selection based on leave-one-out cross-validation*. *IFAC*, 10 (2009), 874-879.
- [6] J. Pohjankukka, T. Pahikkala, P. Nevalainen, & J. Heikkonen, *Estimating the prediction performance of spatial models via spatial k-fold cross-validation*. *Int. J. Geogr. Inf. Syst.*, 10 (2017), 2001-2019.
- [7] T. M. Dutschmann, L. Kinzel, A. ter Laak, & K. Baumann, *Large-scale evaluation of k-fold cross-validation ensembles for uncertainty estimation*. *J. Cheminf.*, 15 (2023), (1), 49.
- [8] X. Zhang, & C. Liu, *Model averaging prediction by k-fold cross-validation*. *J. Econom.*, 235 (2023), 280-301.
- [9] S. Prusty, S. Patnaik, & S.K. Dash, *SKCV: Stratified K-fold cross-validation on ML classifiers for predicting cervical cancer*. *Front. nanotechnol.* (2022).
- [10] P. C. Bürkner, J. Gabry, & A. Vehtari, *Efficient leave-one-out cross-validation for Bayesian non-factorized normal and Student-t models*. *Comput. Stat.*, preprint (2020), 1-19.
- [11] C. Rushing, A. Bulusu, H. I. Hurwitz, A. B. Nixon, *A leave-one-out cross-validation SAS macro for the identification of markers associated with survival*. *Comput. Biol. Med.*, 65 (2015), 123-129.