Computer Science Machine Learning

What is cross validation in machine learning?

This question previously had details. They are now in a comment.

Answer Follow · 14 Request 1

8 Answers

Ishan Shah, Associate at Quantra
Updated Jan 29

In general, we partition the dataset into training and test sets. Then, call the fit method on the training set to build the model and apply the model on the test set to estimate the target value and evaluate the model's performance. The reason why we divided the data into training and test sets was to use the test set to estimate how well the model trained on the training data and how well it would perform on the unseen data.

However, cross-validation is a method that goes beyond evaluating a single model using a single train and test split of the data. It is applied to more subsets created using the training dataset and each of which is used to train and evaluate a separate model. That is, we split our training dataset into k subsets and the ith model will be built on the union of all subsets except the ith. We then test for the performance of the model i on the ith part.

Related Questions

Why do we need to use cross validation in machine learning?

What is cross validation?

What is the difference between validation and cross-validation?

Machine Learning: Can we use Cross Validation instead of test data?

What is generalized cross validation?

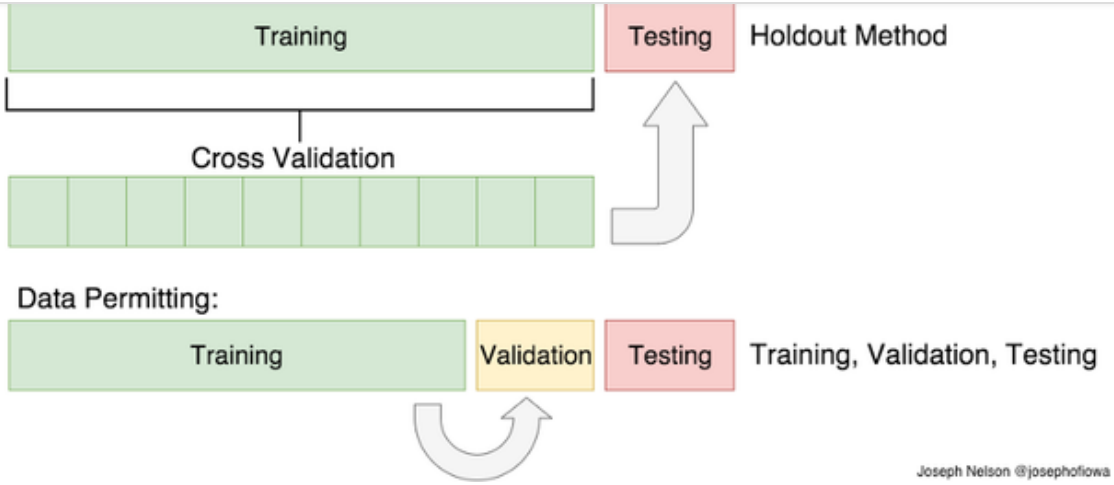
What are the steps of cross validations in machine learning?

How does k-fold cross validation work?

When can cross validation be skipped in a machine learning technique?

+ Ask New Question

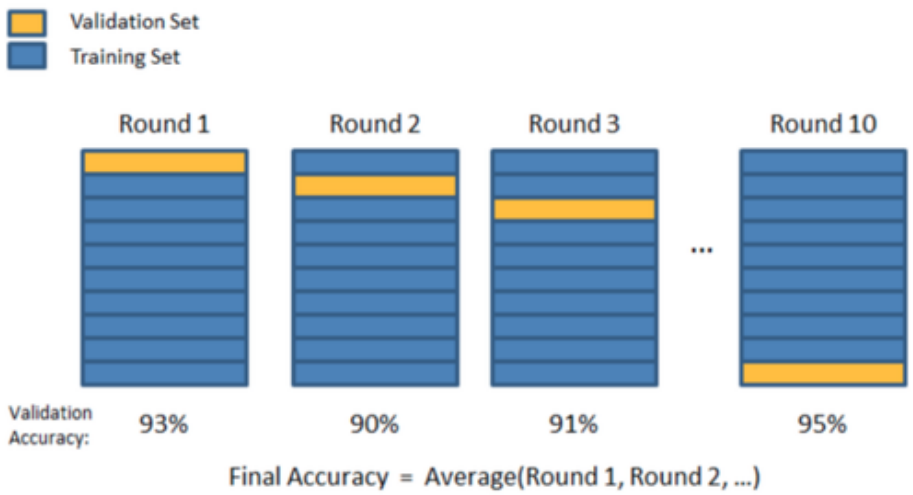
More Related Questions



So why is this better than our original method of a single train and test split?

Well, you may have noticed that by choosing a different length for the train and the test data split, the model performance will vary quite a bit, depending on the specific samples that happen to end up in the training set. Cross-validation gives more stable estimates of how the model is likely to perform on average, instead of relying completely on a single training set.

The most common type of cross-validation technique is the k-fold cross-validation. To do a five-fold cross-validation, the training dataset is partitioned into five parts of equal or close to equal size. Each of these parts is called a "fold". The first model is trained using folds 1 through 4 and evaluated on fold 5. The second model is trained using folds 1, 2, 3, and 5 and evaluated on fold 4, and so on. When this process is done, we have five accuracy values, one for each model. It's typical to then compute the mean and standard deviation of all the accuracy



How do we select the number of folds?

The choice of number of folds must allow the size of each validation partition to be large enough to provide a fair estimate of the model's performance on it and shouldn't be be too small, say 2, such that we don't have enough trained models to evaluate.

Source:

1. Cross Validation in Machine Learning Trading Models
2. Cross-Validation - Amazon Machine Learning

· View Upvoters

Upvote · 11 Share

Add a comment... Recommended All

Frank Gabel, CEO at Machinelearningtutorial.net (2016-present)
Answered Aug 19, 2017

Cross-Validation (in short: CV) has to be put into perspective of splitting your dataset into **train, test** and **validation sets.** Now, CV helps creating multiple "fake" validation sets and averaging their respective errors.

For a graphical explanation, see Feature Engineering in Machine Learning

3.7k Views · View Upvoters

Upvote · 1 Share

Add a comment... Recommended All

Gaurav Chawla, In Quest of Knowledge.....
Answered Nov 14, 2014

I am assuming you know the concept of validation.

I will give a simple example of 10-fold cross validation (CV)
Consider a set of 100 observations.
For first CV, we skip first ten rows for validation set rest 90 used for training. We find MSE for this step.

.

.

.

Similarly for tenth CV we skip last ten rows for validation set and rest 90 used for training. We find MSE for this too.

Now for this 10 fold CV the total MSE is the average of the above MSEs.
We do this because we get dif...(more)

Udit Sabharwal
Answered Feb 13, 2017

Cross-Validation is a statistical method of evaluating and comparing learning algorithms by dividing data into two segments: one used to learn or train a model and the other used to validate the model [1]

Example: Let's say that a trader wants to find a statistical model that attempts to predict the price level of the FTSE100 based on the previous days prices. If he is

Using cross-validation he can estimate the test error of particular statistical learning methods (i.e. their separate predictive perform...(more)

Ndamulelo Banzyme Nemakhavhani, Bsc Electrical Engineering & Machine Learning, University of Cape Town (2016)
Answered Apr 15, 2018

Lets start with the terminology:

1. **Validation** - The process of evaluating thew accuracy of something, i.e. Evaluating mean squared errors(residuals) in supervised regression problems

Background

- Basically the idea of cross validation is a technique used to remove memorizing from the process of machine learning. Remember: Machine learning is all about **teaching computers to generalize**!
- That is if you were to use the same data for training, and validation of your model, you'd simply be teaching your machine to **memorize** - ***Not good!!***

Cross validation approach

- Split your data such that you use a percentag...

Add a comment... Recommended All

Kishore Vasan, recently fascinated by ML
Answered Jul 9, 2016

Cross Validation set is typically 20% of the entire data that you have(60% being the training set and 20% being the test set.)
Cross Validation set provides a better analysis of the theta vector and **chooses a suitable theta** which gives the **lowest cost function**. This theta value so obtained is used in the test set data by x(input) to better predict y(output).
So cross validation basically provides a screening process of the theta variables
I.e **trains the model to better predict values.**

4k Views · View Upvoters

Upvote · 3 Share

Add a comment... Recommended All

Top Stories from Your Feed