Training set

In the context of machine learning, a training set refers to a subset of data used to train a model. The training set is a collection of input-output pairs that the model uses to learn the relationships between inputs and their corresponding outputs. The goal is for the model to generalize from the training data and make accurate predictions on new, unseen data.

Here's a breakdown of the key terms:

Input Data: This is the information provided to the model for training. It could be anything from images and text to numerical data.

Output Data (Labels): For supervised learning, the training set includes not only the input data but also the correct output or label associated with each input. The model learns to map inputs to outputs during the training process.

Training Process: The model iteratively processes the training set, adjusting its internal parameters to minimize the difference between its predictions and the actual labels in the training data. This process involves optimization algorithms like gradient descent.

Generalization: The ultimate aim of training is to enable the model to generalize well to new, unseen data. A well-trained model should be able to make accurate predictions on data it hasn't encountered before.

Typically, a dataset is divided into three parts: training set, validation set, and test set. The training set is used to train the model, the validation set is used to tune hyperparameters and prevent overfitting, and the test set is used to evaluate the model's performance on unseen data.

It's important to have a representative and diverse training set to ensure that the model learns a broad range of patterns and can generalize effectively. The quality and size of the training set play a crucial role in the performance of the machine learning model.

From:
https://neurosurgerywiki.com/wiki/ - **Neurosurgery Wiki**

Permanent link:
https://neurosurgerywiki.com/wiki/doku.php?id=training_set

Last update: **2024/06/07 02:48**