In-database machine learning

Last Updated: 2022-10-20

6

With machine learning, you can create a statistical model using data from your Db2® database. Machine learning is a powerful solution for solving complex problems.

Note: This feature is available starting from Db2 version 11.5.4.

Use cases for machine learning solutions include:

- Problems for which existing solutions require long lists of rules – a machine learning algorithm can often simplify the problem and outperform this traditional approach.
- Problems for which a traditional approach does not provide a satisfactory solution.
- Fluctuating environments.
- Analyzing large amounts of complex data to gain insights.
- Learning and exploiting "latent" or "hidden" features from the data, otherwise invisible to humans and traditional statistical methods.

Generally, machine learning is divided into three categories:

Supervised learning
Supervised learning is performed using a "ground truth" where you have prior knowledge of what the output values will be. These ground truth values are called "labels" or "targets". For example, when attempting to predict credit card fraud, the training data would include past transactions which are definitively fraudulent (the target).

Reinforcement learning
With reinforcement learning, a machine or an "agent" observes its environment, takes "actions", receives "rewards", and learns decides what further actions to preform based on what will yield the maximum "reward". Reinforcement learning is different from supervised leaning, as it has a prior knowledge the "reward" concept in relation to particular "actions". Additionally, this method does not rely on labeled data.

Unsupervised learning
With unsupervised learning, the training data does not include any labels. Having no labeled outputs, the model needs to infer the natural structure that the data points have in common. As such, unsupervised learning tasks typically involve finding underlying structure or patterns in the data that otherwise may not be obvious.

Typical machine learning tasks include:

Classification
Classifying instances into one of multiple categories. For example, classifying emails as "spam" or "not spam".

Regression
Predicting a target numerical value. For example, estimating the price of a house.

Clustering
Detecting groups within the data. For example, detecting groups of similar visitors to a website

Vast amounts of data are often required to produce a robust and accurate predictive model. With machine learning for Db2, a user can create a machine learning model without moving data from Db2. This enhances security, as data is not moved at any time from the secure database. It also enhances speed, as there is no data transfer cost.

- **Prerequisites for machine learning in Db2**
  You must meet the following requirements to use the machine learning functionality in Db2.
- **Machine learning stored procedures**
  Machine learning stored procedures execute SQL queries in the Db2 database, performing common machine learning tasks such as data transformation, data processing, model building, and model evaluation. You can use these stored procedures to build a complete machine learning pipeline (data exploration, data transformation, model selection, model training, model evaluation, and model deployment) entirely within Db2 without transferring data from Db2 to another system. You can build models for classification and regression. The stored procedures are listed in alphabetical order.

Related reference

→ Troubleshooting analytic stored procedures