Postgres and the Artificial Intelligence Landscape

Bruce Momjian

This presentation explains how to do machine learning inside the Postgres database.

https://momjian.us/presentations

Last updated: October 2024

Outline

1. What is artificial intelligence?
2. Machine learning and deep learning
3. Demonstration using Postgres
4. Hardware/software efficiency
5. Tasks
6. Why use a database?

1. What is Artificial Intelligence?

Machines that mimic “cognitive” functions that humans associate with the human mind, such as “learning” and “problem solving”.

https://en.wikipedia.org/wiki/Artificial_intelligence

What is Artificial about Artificial Intelligence?

If only the physical world exists, then human intelligence only differs from machine intelligence because it has not naturally developed. It hence differs only in how it is created. Human free will becomes an illusion.

https://www.theatlantic.com/magazine/archive/2016/06/theres-no-such-thing-as-free-will/480750/

History of Artificial Intelligence (AI)

- Pre-computer philosophy
- Robotics
- Turing test
- Expert systems
- AI winter
- Machine/deep learning
- Generative AI: output text, images, videos
- Like fusion energy, it is always ten years away

https://en.wikipedia.org/wiki/History_of_artificial_intelligence

Artificial Intelligence Naming

”In a certain sense I think that artificial intelligence is a bad name for what it is we’re doing here. As soon as you utter the words ’artificial intelligence’ to an intelligent human being, they start making associations about their own intelligence, about what’s easy and hard for them, and they superimpose those expectations onto these software systems.”

Kevin Scott, chief technology officer of Microsoft.

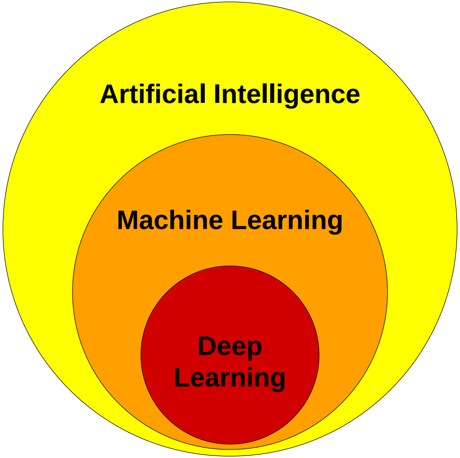
https://www.wsj.com/articles/why-artificial-intelligence-isnt-intelligent-11627704050?st=6qiqn8g4h5kld5u&reflink=desktopwebshare_permalink

Artificial Intelligence Reality

"AI algorithms are just math. And one of math's functions is to simplify the world so our brains can tackle its otherwise dizzying complexity. The software we call AI is just another way to arrive at complicated mathematical functions that help us do that."

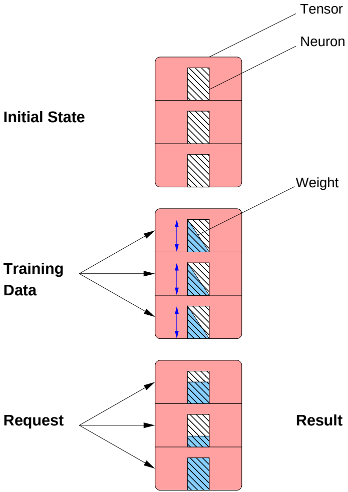
Kevin Scott, chief technology officer of Microsoft.

https://www.wsj.com/articles/why-artificial-intelligence-isnt-intelligent-11627704050?st=6qiqn8g4h5kld5u&reflink=desktopwebshare_permalink



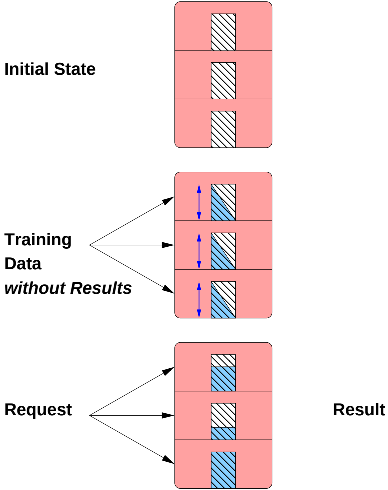
The most comprehensive video I have seen about machine learning is at https://www.youtube.com/watch?v=r0Ogt-q956I.

Machine Learning



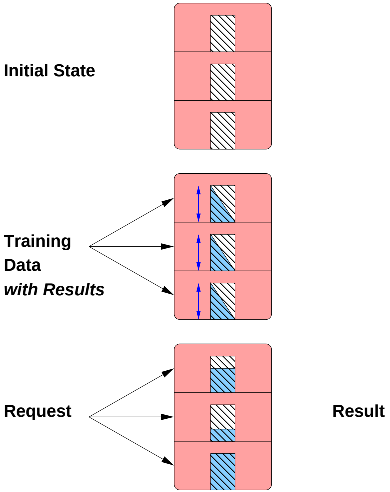
https://www.verypossible.com/insights/machine-learning-algorithms-what-is-a-neural-network

Unsupervised Machine Learning

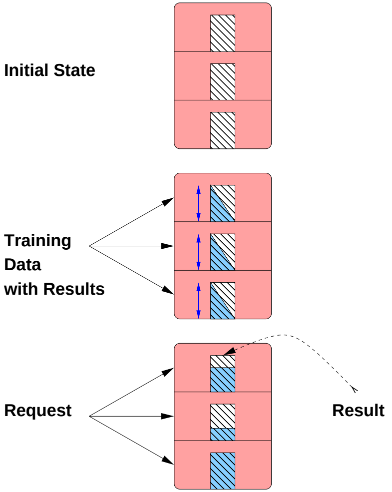


https://medium.com/@machadogj/ml-basics-supervised-unsupervised-and-reinforcement-learning-b18108487c5a

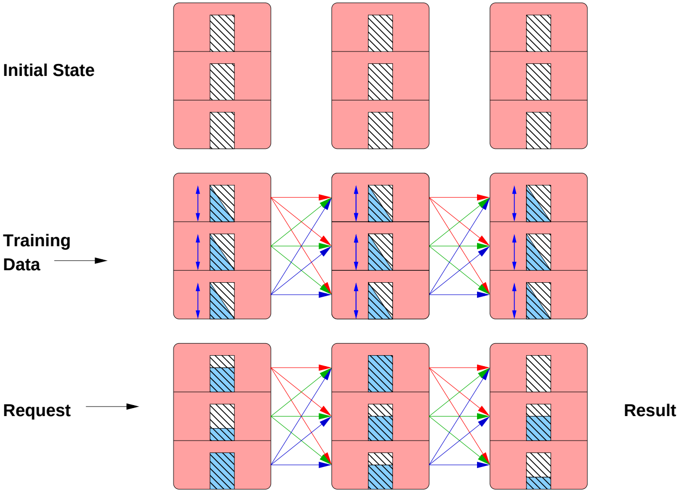
Supervised Machine Learning



Reinforcement Machine Learning



Deep Learning



https://www.zendesk.com/blog/machine-learning-and-deep-learning/

3. Demonstration Using Postgres: Does an Integer Have Non-Leading Zeros?

- 31903 is true
- 82392 is false

Install PL/Perl

```
CREATE EXTENSION IF NOT EXISTS plperl;
```

All queries in this presentation can be downloaded from https://momjian.us/main/writings/pgsql/landscape.sql.

Generate Tensor

```
CREATE OR REPLACE FUNCTION generate_tensor(value INTEGER)
RETURNS BOOLEAN[] AS $$
        my $value = shift;
        my @tensor = (
                # this many digits or more?
                (map { length($value) >= $_ } 1..10),
                # divisible by 10?
                $value % 10 == 0,
        );
        # map to t/f
        grep { $_ = ($_ ? 't' : 'f') } @tensor;
        return encode_typed_literal(\@tensor, 'boolean[]');
$$ LANGUAGE plperl STRICT;
```

The tensor has 11 neurons.

Create and Populate Input Layer with Training Data

```
CREATE TABLE training_set(value INTEGER, training_output BOOLEAN,
                          tensor BOOLEAN[]);
WITH randint (value) AS
(
        SELECT (random() * (10 ^ (random() * 8 + 1)::integer))::integer
        FROM generate_series(1, 10000)
)
INSERT INTO training_set SELECT value, value::text LIKE '%0%',
                                generate_tensor(value)
FROM randint;
```

Input Layer

```
SELECT * FROM training_set LIMIT 10;
   value    | training_output |         tensor
-----------+-----------------+-------------------------
   8360692 | t               | {t,t,t,t,t,t,t,f,f,f,f}
  58297366 | f               | {t,t,t,t,t,t,t,t,f,f,f}
    569005 | t               | {t,t,t,t,t,t,f,f,f,f,f}
       236 | f               | {t,t,t,f,f,f,f,f,f,f,f}
        43 | f               | {t,t,f,f,f,f,f,f,f,f,f}
        35 | f               | {t,t,f,f,f,f,f,f,f,f,f}
    610510 | t               | {t,t,t,t,t,t,f,f,f,f,t}
 638484259 | f               | {t,t,t,t,t,t,t,t,t,f,f}
      5983 | f               | {t,t,t,t,f,f,f,f,f,f,f}
        34 | f               | {t,t,f,f,f,f,f,f,f,f,f}
```

Generate Weights for Tensor

```
-- Runs the supplied query and generates weights
CREATE OR REPLACE FUNCTION generate_weight(query TEXT, desired_output BOOLEAN)
RETURNS REAL[] AS $$
        my $rv = spi_exec_query(shift);
        my $status = $rv->{status};
        my $nrows = $rv->{processed};
        my $desired_output = shift;
        my @success_neurons = ();
        my @desired_neurons = ();
        my $desired_input = 0;
```

Generate Weights for Tensor

```
foreach my $rn (0 .. $nrows - 1) {
            my $row = $rv->{rows}[$rn];
            my $tensor = $row->{(sort keys %$row)[0]};
            my $training_output = $row->{(sort keys %$row)[1]};
            # only process training rows that match our desired output
            foreach my $neuron (0 .. $#$tensor)
            {
                    $success_neurons[$neuron] //= 0;
                    $desired_neurons[$neuron] //= 0;
                    # Neuron value matches desired output value;  does
                    # the value match the desired output?
                    if ($tensor->[$neuron] eq $desired_output)
                    {
                            # Prediction success/failures that match our
                            # desired output.
                            $success_neurons[$neuron]++
                                    if ($training_output eq $desired_output);
                            $desired_neurons[$neuron]++;
                    }
            }
            $desired_input++ if ($training_output eq $desired_output);
    }
```

Generate Weights for Tensor

```
        my @weight = ();
        my $sum = 0;

    # compute percentage of tests that matched requested outcome
        foreach my $neuron (0 .. $#success_neurons) {
                $weight[$neuron] = $desired_neurons[$neuron] != 0 ?
                        $success_neurons[$neuron] / $desired_neurons[$neuron] :
                        0;
                $sum += $weight[$neuron];
        }

        # balance weights so they total the observed probability;
        # this prevents an overly-predictive output value from skewing
        # the results.
        foreach my $neuron (0 .. $#weight) {
                $weight[$neuron] = ($weight[$neuron] / $sum) *
                                   ($desired_input / $nrows);
        }
        return encode_typed_literal(\@weight, 'real[]');
$$ LANGUAGE plperl STRICT;
```

Create Tensor_Mask

```
# Return weights where our neuron value matches the desired output
CREATE OR REPLACE FUNCTION tensor_mask(tensor BOOLEAN[], weight REAL[],
                                          desired_output BOOLEAN)
RETURNS REAL[] AS $$
        my $tensor = shift;
        my $weight = shift;
        my $desired_output = shift;
        my @result = ();

       elog(ERROR, 'tensor and weight lengths differ')
                if ($#$tensor != $#$weight);
        foreach my $i (0 .. $#$tensor) {
                push(@result,
                        ($tensor->[$i] eq $desired_output) ?
                        $weight->[$i] : 0);
        }
        return encode_typed_literal(\@result, 'real[]');
$$ LANGUAGE plperl STRICT;
```

Create Sum_Weight

```
CREATE OR REPLACE FUNCTION sum_weight(weight REAL[])
RETURNS REAL AS $$
        my $weight = shift;
        my $sum = 0;
        # sum weights
        foreach my $i (0 .. $#$weight) {
                $sum += $weight->[$i];
        }
        return encode_typed_literal($sum, 'real');
$$ LANGUAGE plperl STRICT;
```

Create Soft_Max

```
# Normalize the values so the probabilities total one
CREATE OR REPLACE FUNCTION softmax(val1 REAL, val2 REAL)
RETURNS REAL[] AS $$
        my $val1 = shift;
        my $val2 = shift;
        my $sum = exp($val1) + exp($val2);
        # What percentage is each of the total?
        my @result = (
                exp($val1) / $sum,
                exp($val2) / $sum,
        );
        return encode_typed_literal(\@result, 'real[]');
$$ LANGUAGE plperl STRICT;
```

Uses the exponential function (e^x)

Store Weights

```
CREATE TABLE tensor_weight_true AS
SELECT generate_weight('SELECT tensor AS x1, training_output AS x2
                        FROM training_set', true) AS weight;

CREATE TABLE tensor_weight_false AS
SELECT generate_weight('SELECT tensor AS x1, training_output AS x2
                        FROM training_set', false) AS weight;
```

Stored Weights

```
SELECT * FROM tensor_weight_true;
                            weight
----------------------------------------------------------------
 {0.021163348,0.022521615,0.024933573,0.027230926,0.02884723, \
  0.030753216,0.032510195,0.033661984,0.036457982,0,0.065419935}

SELECT * FROM tensor_weight_false;
                           weight
--------------------------------------------------------------
 {0,0.080344856,0.07735135,0.07436457,0.07038565,0.067453355, \
  0.06445343,0.061397385,0.05898541,0.057765257,0.06399873}
```

Test 100

```
WITH test_set (checkval) AS
(
        SELECT 100
)
SELECT softmax(
        sum_weight(
                tensor_mask(
                        generate_tensor(checkval),
                        tensor_weight_true.weight,
                        true)),
        sum_weight(
                tensor_mask(
                        generate_tensor(checkval),
                        tensor_weight_false.weight,
                        false))
)
FROM test_set, tensor_weight_true, tensor_weight_false;
         softmax
-------------------------
 {0.42048895,0.57951105}
```

Test 101

```
WITH test_set (checkval) AS
(
        SELECT 101
)
SELECT softmax(
        sum_weight(
                tensor_mask(
                        generate_tensor(checkval),
                        tensor_weight_true.weight,
                        true)),
        sum_weight(
                tensor_mask(
                        generate_tensor(checkval),
                        tensor_weight_false.weight,
                        false))
)
FROM test_set, tensor_weight_true, tensor_weight_false;
         softmax
------------------------
 {0.3893167,0.61068326}
```

Test 487234987

```
WITH test_set (checkval) AS
(
        SELECT 487234987
)
SELECT softmax(
        sum_weight(
                tensor_mask(
                        generate_tensor(checkval),
                        tensor_weight_true.weight,
                        true)),
        sum_weight(
                tensor_mask(
                        generate_tensor(checkval),
                        tensor_weight_false.weight,
                        false))
)
FROM test_set, tensor_weight_true, tensor_weight_false;
         softmax
------------------------
 {0.5340263,0.46597365}
```

Test One Thousand Values

```
WITH test_set (checkval) AS
(
        SELECT (random() * (10 ^ (random() * 8 + 1)::integer))::integer
        FROM generate_series(1, 1000)
),
```

Second Table Expression

```
ai (checkval, output_layer) AS
(
        SELECT checkval, softmax(
                sum_weight(tensor_mask(generate_tensor(checkval),
                           tensor_weight_true.weight, true)),
                sum_weight(tensor_mask(generate_tensor(checkval),
                           tensor_weight_false.weight, false))
        )
        FROM test_set, tensor_weight_true, tensor_weight_false
),
```

Third Table Expression

```
analysis (checkval, cmp_bool, output_layer, accuracy) AS
(
        SELECT checkval, checkval::text LIKE '%0%', output_layer,
                CASE checkval::text LIKE '%0%'
                -- higher/lower than random chance
                WHEN true THEN output_layer[1] - 0.5
                ELSE output_layer[2] - 0.5
                END
        FROM ai
)
```

Final Table Expression

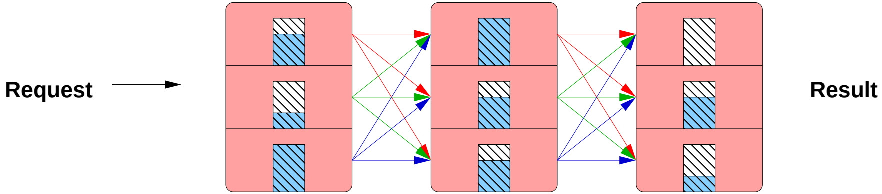
```
SELECT * FROM analysis
UNION ALL
SELECT NULL, NULL, NULL, AVG(accuracy)
FROM analysis
UNION ALL
SELECT NULL, NULL, NULL, SUM(CASE WHEN accuracy > 0 THEN 1 END)::REAL/COUNT(*)
FROM analysis;
 checkval  | cmp_bool |      output_layer       |        accuracy
-----------+----------+-------------------------+-----------------------
  35173962 | f        | {0.5102168,0.48978326}  | -0.0102167427539825
     27023 | t        | {0.4379818,0.5620182}   |  -0.0620181858539581
   9669661 | f        | {0.48645663,0.51354337} |  0.0135433673858642
   6348988 | f        | {0.48645663,0.51354337} |  0.0135433673858642
      7736 | f        | {0.41372445,0.5862756}  |   0.0862755775451660
    910139 | t        | {0.4622842,0.5377158}   | -0.0377157926559448
      8642 | f        | {0.41372445,0.5862756}  |   0.0862755775451660
        98 | f        | {0.3652915,0.6347085}   |   0.134708523750305
         6 | f        | {0.34178793,0.65821207} |    0.158212065696716
       962 | f        | {0.3893167,0.61068326}  |   0.110683262348175
…
    (null) | (null)   | (null)                  |  0.0496436570882797
    (null) | (null)   | (null)                  |                0.722
```

4. Hardware/Software Efficiency: Software

- Client-side
  - Matlab
  - Scikit, https://kb.objectrocket.com/postgresql/machine-learning-with-python-and-postgres-1114
  - Tensorflow
  - PyTorch
  - Weka
- Server-side
  - PL/Python with the above libraries
  - MADlib, https://www.youtube.com/watch?v=uLW5By66Lf0
  - Scikit, https://www.cybertec-postgresql.com/en/machine-learning-in-postgresql-part-1-kmeans-clustering/
  - pgvector, https://github.com/pgvector/pgvector

https://www.youtube.com/watch?v=gTqVU_bhkmg

GPU



- Tensors can have millions of neurons
- Deep learning can use thousands of tensor layers
- Every neuron passes its data to every neuron in the next layer
- This requires a lot of repetitive calculations
- GPUs are designed to efficiently perform simultaneous repetitive computations
- PG-Strom adds GPU acceleration to Postgres

5. Tasks

- Chess (Deep Blue)
- Jeopardy (Watson)
- Voice recognition (Siri)
- Search (Google)
- Recommendations (Netflix)
- Image detection
- Weather forecasting (NOAA)

Fraud Detection Example

Choose attributes:

- Charge amount
- Magnetic swipe, chip, pin, online charge
- Vendor distance from chargee billing address
- Distance from last chargee charge
- Vendor country
- Previous charges to this vendor for chargee
- Previous fraudulent charges by vendor

https://www.youtube.com/watch?v=ztVcNMg7TFs

Fraud Detection Steps

1. Choose attributes
2. Create machine learning neurons for each attribute
3. Create training data, with the required attributes of each transactions and its outcome, i.e., valid or fraudulent
4. Feed the training data into the machine to set the weights of each neuron, based on how much the neuron's attribute predicts the validity or fraudulence of transactions
5. Start feeding real data into the machine and get results
6. Feed correct and incorrect results back into the neurons to improve the accuracy of the weights, and to adjust for changes in the environment

6. Why Use a Database?

- Machine learning requires a lot of data
- Most of your data is in your database
- Why not do machine learning where your data is, in a database?

Advantages of doing Machine Learning in a Database?

- Use previous activity as training data
- Have seamless access to all your current data
- Take immediate action on AI results, e.g., commit transaction only if likely non-fraudulent
- AI can benefit from database transactions, concurrency, backup
- Other benefits include complex data types, full text search, GIS, indexing
- Postgres can do GPU-based computations inside the database (https://momjian.us/main/blogs/pgblog/2020.html#June_29_2020)

General Artificial Intelligence Uses by Databases?

- User applications not already covered
  - human language queries and results (large language model)
  - retrieval-augmented generation (RAG)
- Performance adjustments
  - optimizer plans
  - index creation/destruction
  - database settings
  - resource usage
- Alerting
  - malicious activity
  - resource exhaustion

Conclusion

https://momjian.us/presentations

https://www.flickr.com/photos/corneveaux/