In-Database Machine Learning
Using Gradient Descent and Tensor Algebra

Maximilian Schüle, Frédéric Simonis, Thomas Heyenbrock, Alfons Kemper, Stephan Günnemann, Thomas Neumann
{schuele, simonis, kemper, guennemann, neumann}@in.tum.de

HyPer + Tensors + Gradient Descent

**Machine Learning**: Data in tensors and a loss function

- Operator for gradient descent:
  - Gradient needed for gradient descent: automatic differentiation necessary for arbitrary loss functions
  - Integration in relational algebra
  - Representation of a loss function

\[ \lambda(R,S)(R.a + S.x + R.b - S.y)^2 \]

- Tensors: datatype with algebra
- Optimisation problems solvable in the core of database systems

**Materializing**
- Allows any optimization method
  - Tuples need to be materialized

**Pipelined**
- No materialization required
  - Iterations must be precompiled
  - Little performance gains

**Combined**
- Precomputes weights in pipelines

**Evaluation**

- Linear Regression
- Multiple Linear Regression
- Logistic Regression
- k-Means