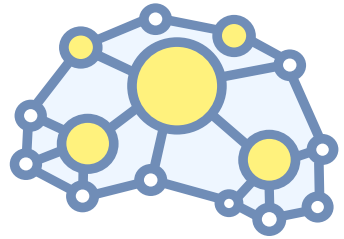
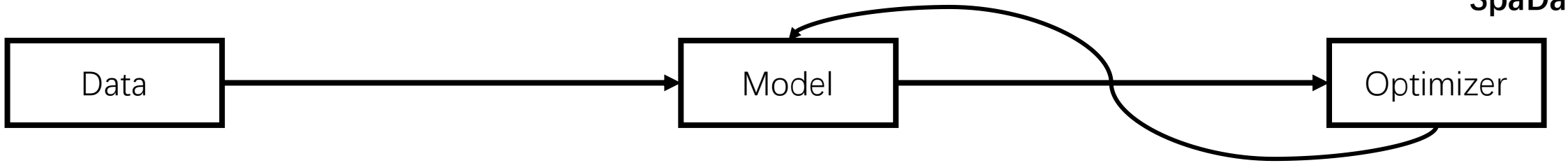
SpaDaFlow: Spatio-Temporal Data Flow



SpaDaFlow



a. Cross-Sectional Data

b. Time Series Data

c. Pooled Cross Sections Data

d. Panel Data

e. Spatial Panel Data

a. One-Stage Model

Traditional statistical models and machine learning models

b. Two-Stage Model:

1. Construct the Spatial Weight Matrix
2. Construct the Structural Model

a. Normal Equation Strategy

b. Greedy Algorithm Strategy

1. Non-Gradient based method
2. Gradient based method



CuPy

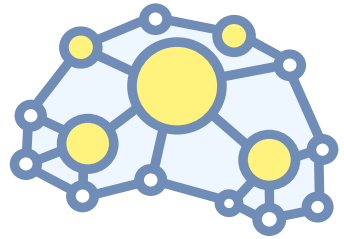



HyperLearn

PyTorch

TensorFlow

SpaDaFlow: Spatio-Temporal Data Flow



SpaDaFlow

Data

Model

Optimizer

Voice Data

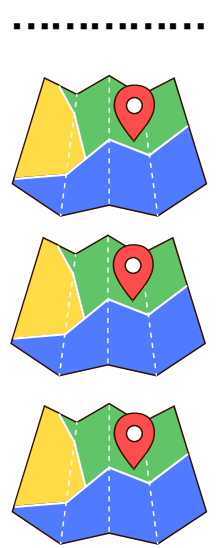
Video Data

Graph Data

Feature-List Data

Picture Data

Context Data



T+1

T

T-1

Spatial Panel Data

CuPy NumPy pandas SciPy SQLite scikit learn HyperLearn PyTorch TensorFlow

SpaDaFlow: Spatio-Temporal Data Flow

SpaDaFlow

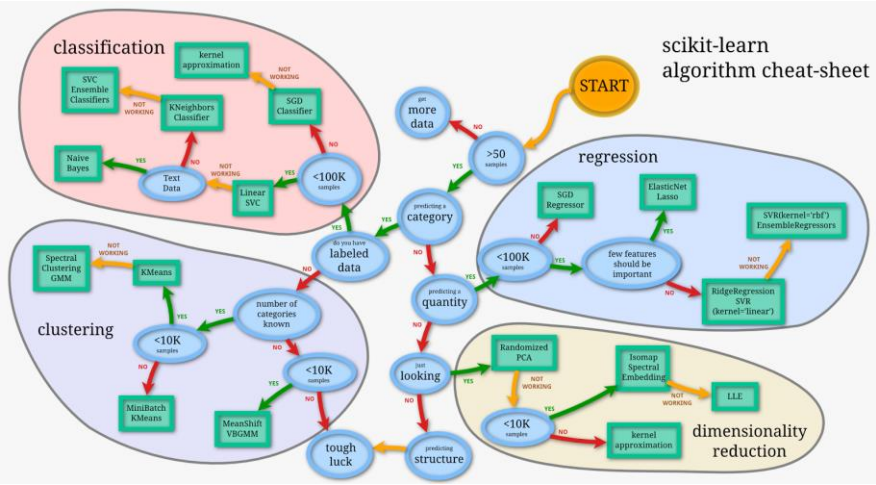
Data

Model

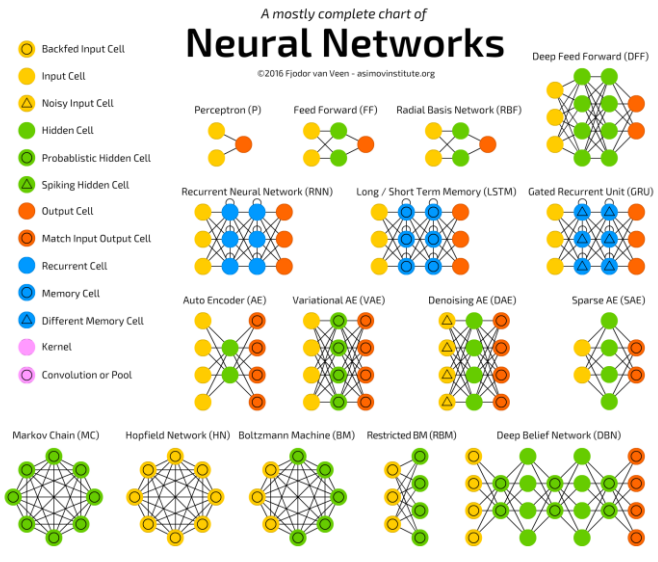
Optimizer

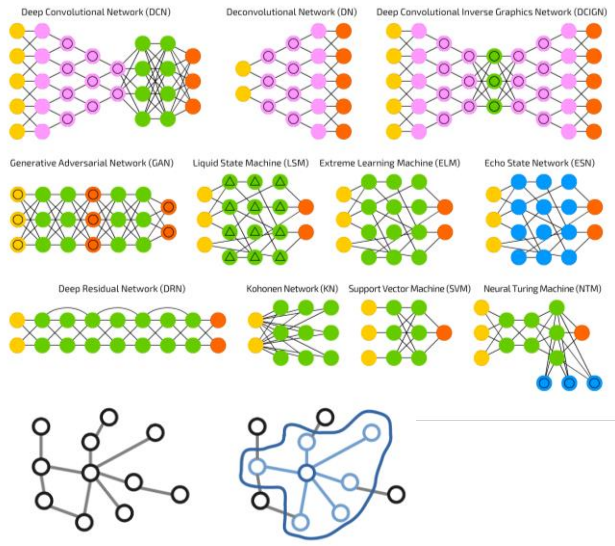
One-Stage Model

Traditional statistical models



Deep learning models







CuPy

NumPy



SciPy

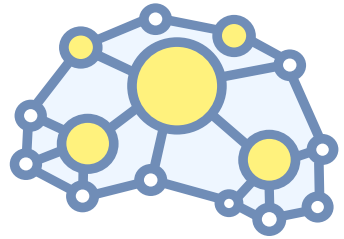
SQLite


HyperLearn

PyTorch

TensorFlow

SpaDaFlow: Spatio-Temporal Data Flow



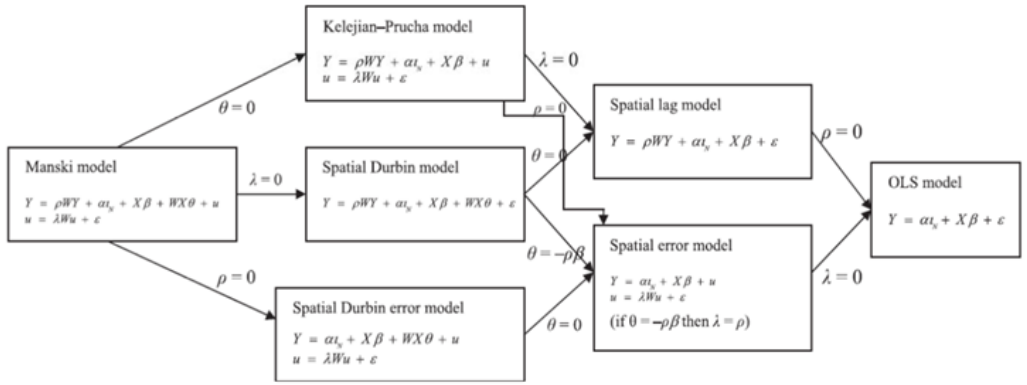
SpaDaFlow

Data | **Model** | Optimizer

Two-Stage Model (1)

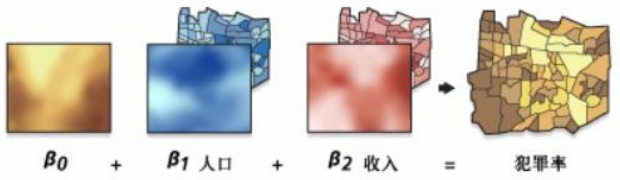
The idea about the two-stage model mainly comes from **spatial econometrics** and **spatial statistics**

Tobler's First Law of Geography
(Spatial Autoregression Problem)

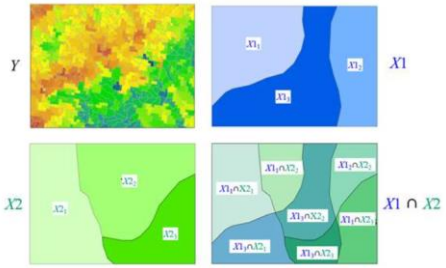


Goodhild's Second Law of Geography
(Spatial Heterogeneity Problem)

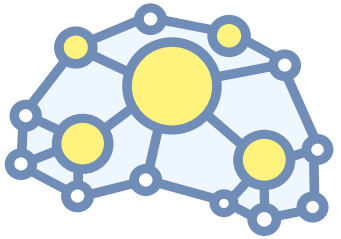
Local Solution:
Geographic Weighted Regression (GWR)



Global Solution:
GeoDetector



CuPy NumPy pandas SciPy SQLite scikit learn HyperLearn PyTorch TensorFlow



SpaDaFlow

Data | Model | Optimizer

Two-Stage Model (2)

All the above solutions can be summarized into the following **two steps**.

1. Construct the Spatial Weight Matrix

Traditional method:

a. Weights Based on Boundaries

b. Weights Based on Distance

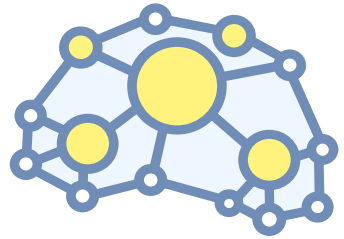
Existing Problem:

The relationship which cannot be described by the linear distance.

Our method in SpaDaFlow:

Using the **machine learning method** to generate complex - relationship-related Spatial Weight Matrix and considering the **interpretability** of this method.

SpaDaFlow: Spatio-Temporal Data Flow



SpaDaFlow

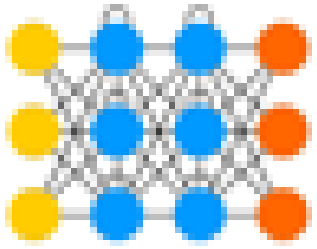
Data

Model

Optimizer

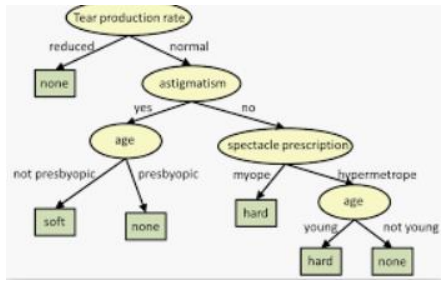
Two-Stage Model (3)

2. Construct the Structural Model

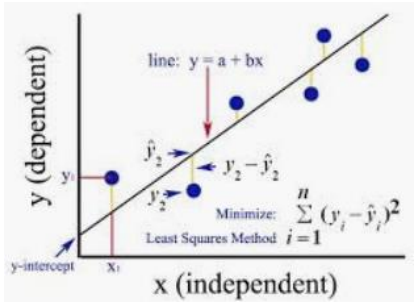


Neural Networks

.....



Tree Model

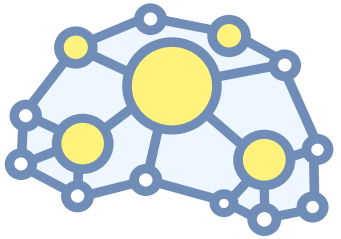


Linear Model

Low **High**

Interpretability

SpaDaFlow: Spatio-Temporal Data Flow



SpaDaFlow

Data

Model

Optimizer

a. Normal Equation Strategy

Normal Equation
Maximum Likelihood Estimate
……

b. Greedy Algorithm Strategy

1. Non-Gradient based method
MCMC、PSO、GA、SA

2. Gradient based method
GD、SGD、Adam……
Newton's method



CuPy

NumPy

pandas

SciPy

SQLite


HyperLearn
PyTorch

TensorFlow

SpaDaFlow: Spatio-Temporal Data Flow

Case: Industrial Data (620 simples, 21 features inside)

Early Stop: 679 Epoch

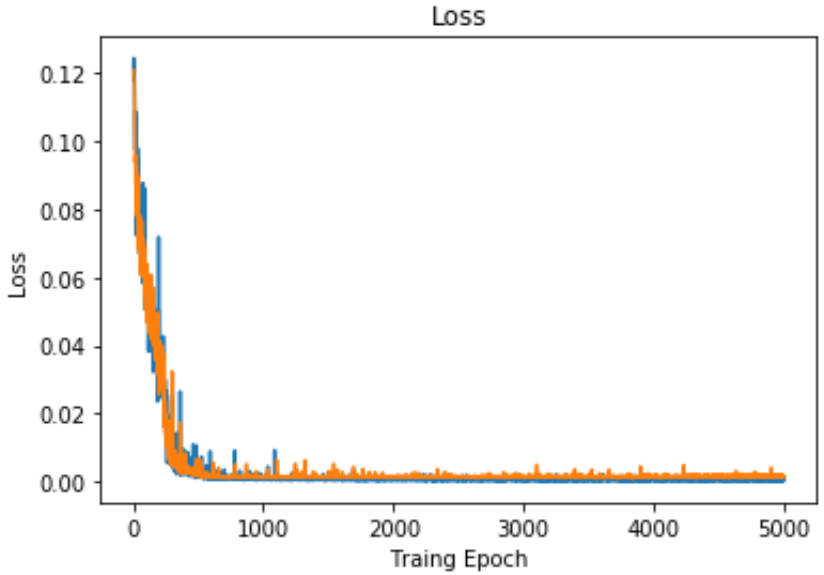
Best Model (in 679 Epoch)

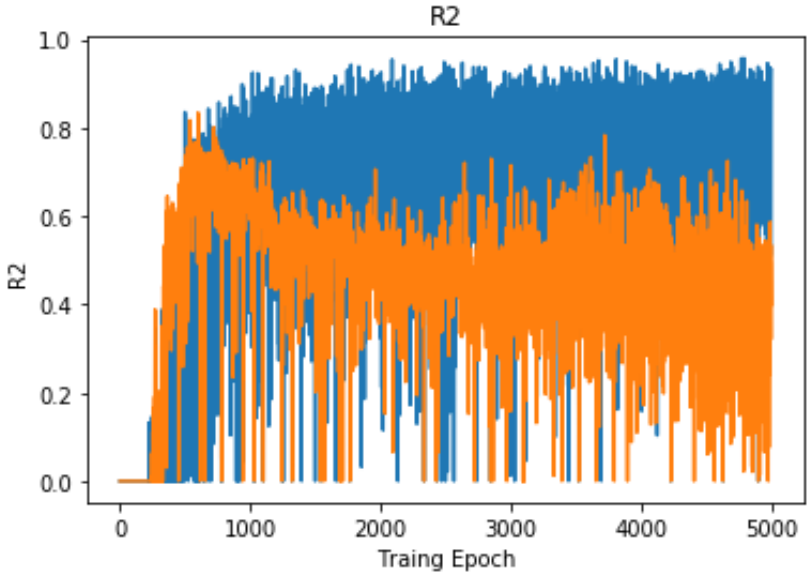
Train Loss: 0.0005

Test Loss: 0.0012

Train R2: 0.8434583775428112

Test R2: 0.7373595835912443





SpaDaFlow

CuPy NumPy pandas SciPy SQLite scikit learn HyperLearn PyTorch TensorFlow