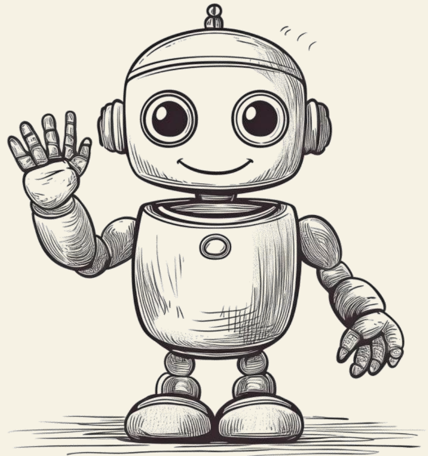
Neural ODE Tutorial

Can robots do differential equations?

Moritz Laber
Friday, 20th June 2025 at SFI CSSS2025

laber.m@northeastern.edu



The Good Old Days of Mathematical Modelling

Some people just talk to much…don’t you think so, too?

The Good Old Days of Mathematical Modelling

Some people just talk to much…don't you think so, too?

There are way so many rumors going around at this summer school.

The Good Old Days of Mathematical Modelling

Some people just talk to much…don't you think so, too?

There are way so many rumors going around at this summer school.

Can we find a quantitative, mathematical description of this phenomenon?

The Good Old Days of Mathematical Modelling

Some people just talk to much…don't you think so, too?

There are way so many rumors going around at this summer school.

Can we find a quantitative, mathematical description of this phenomenon?

Of course! Let's use *differential equations*!

The Good Old Days of Mathematical Modelling

Maki-Thompson Rumor Model

I **informed individuals**: know the rumor, spread it..
S **susceptible individuals**: don't know the rumor but…would like to.
R **removed individuals**: know the rumor, don't care to spread it.

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta is - \mu i^2 - \mu ir$$

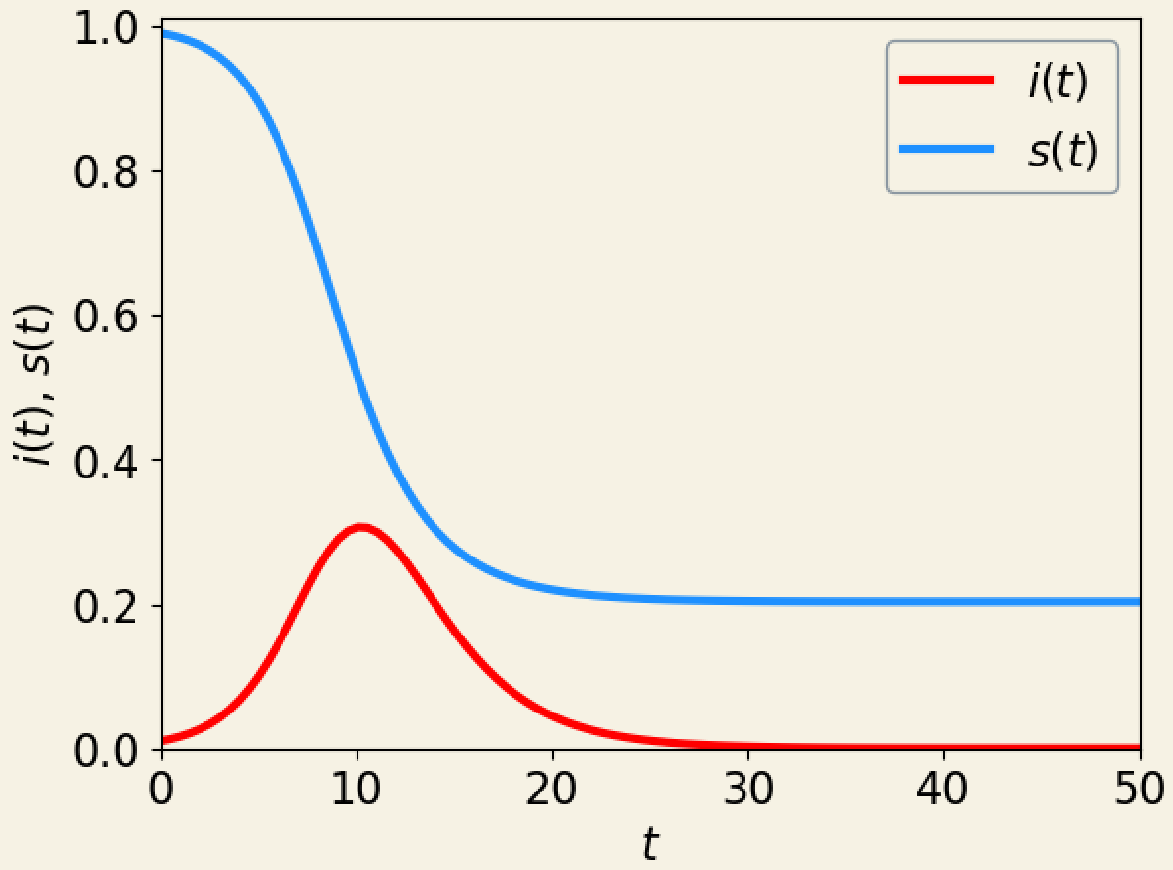
$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta is$$

$$r = 1 - s - i$$

The Good Old Days of Mathematical Modelling

Numerical Solution

- Fix an initial condition: $(i_0, s_0, 1 - s_0 - i_0)$.
- Integrate the differential equation



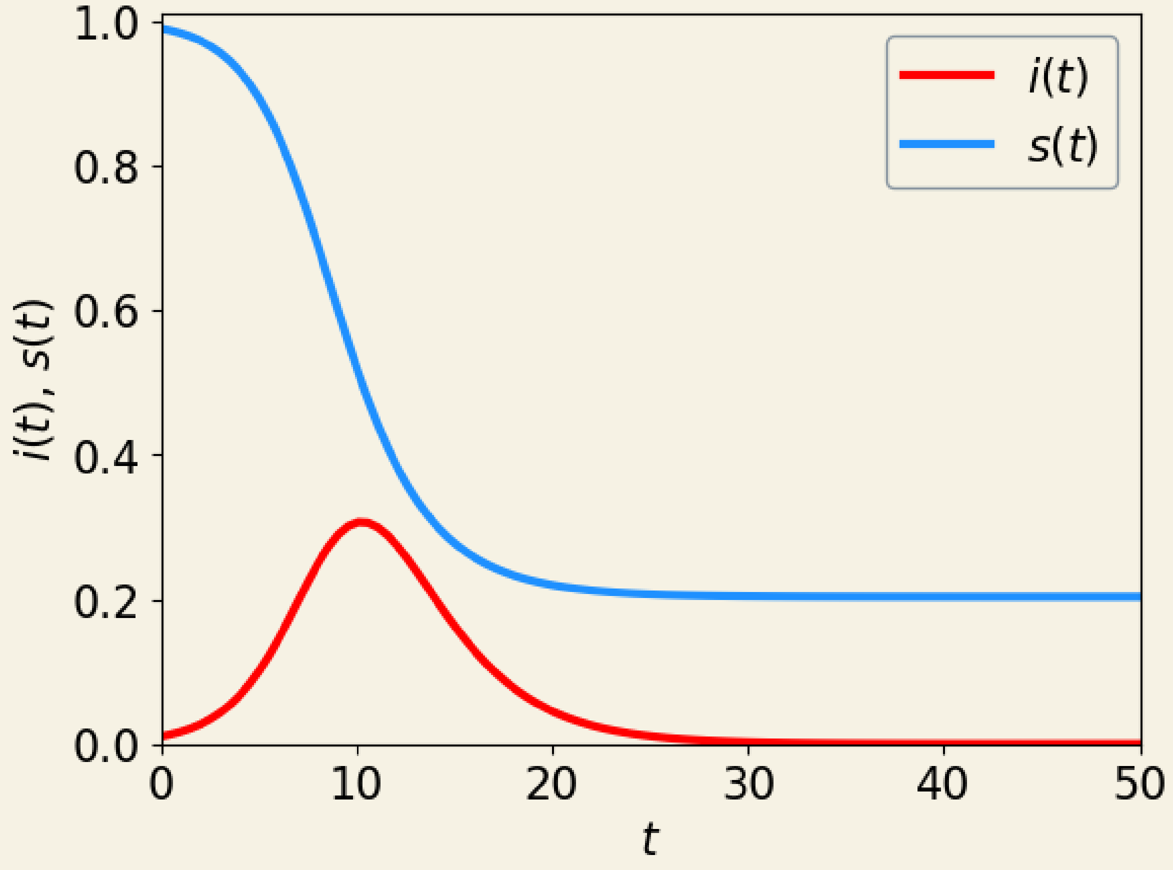
The Good Old Days of Mathematical Modelling

Numerical Solution

- Fix an initial condition: $(i_0, s_0, 1 - s_0 - i_0)$.
- Integrate the differential equation

Analytical Treatment

- Fixed Points and Stability Analysis
- Steady States



A Modern, Data-Driven Twist on an Old Story

What did we do well?

- Interpretable: All terms have clear meaning.
- Robust: Captures main trends.

A Modern, Data-Driven Twist on an Old Story

What did we do well?

- Interpretable: All terms have clear meaning.
- Robust: Captures main trends.

What are potential problems?

- Simplistic: Can‘t capture fine details.
- Relies on prior understanding.

A Modern, Data-Driven Twist on an Old Story

What did we do well?

- Interpretable: All terms have clear meaning.
- Robust: Captures main trends.

What are potential problems?

- Simplistic: Can‘t capture fine details.
- Relies on prior understanding.

➔ **Idea:** Learn the form of the differential equation from data!

A Modern, Data-Driven Twist on an Old Story

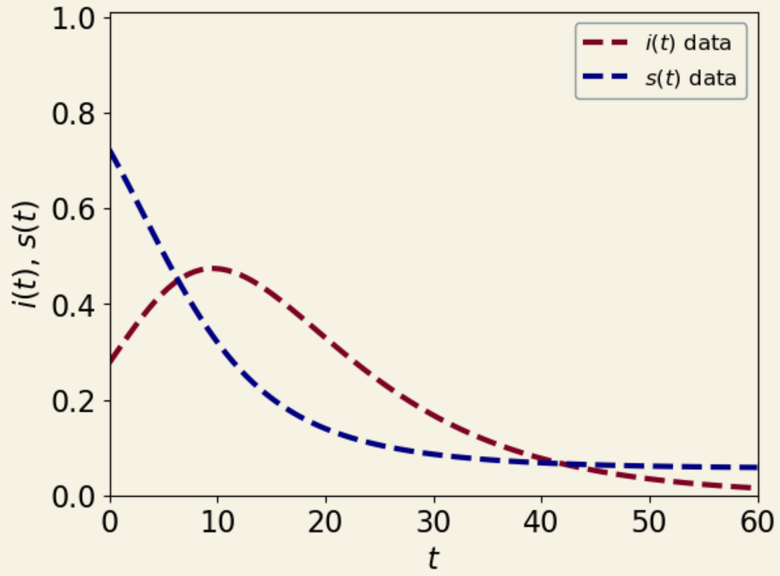
What did we do well?

- Interpretable: All terms have clear meaning.
- Robust: Captures main trends.

What are potential problems?

- Simplistic: Can‘t capture fine details.
- Relies on prior understanding.

➔ **Idea:** Learn the form of the differential equation from data!



$$\frac{\mathrm{d}i}{\mathrm{d}t} = f(i, s)$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = g(i, s)$$

A Modern, Data-Driven Twist on an Old Story

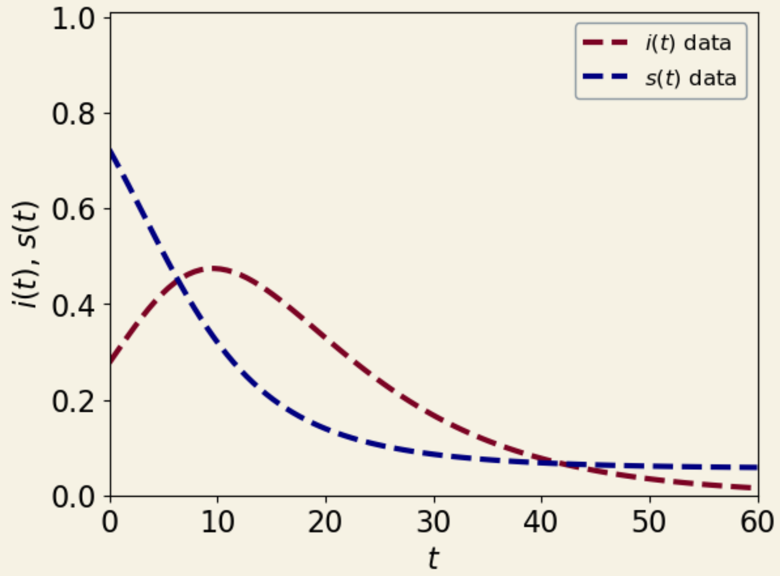
What did we do well?

- Interpretable: All terms have clear meaning.
- Robust: Captures main trends.

What are potential problems?

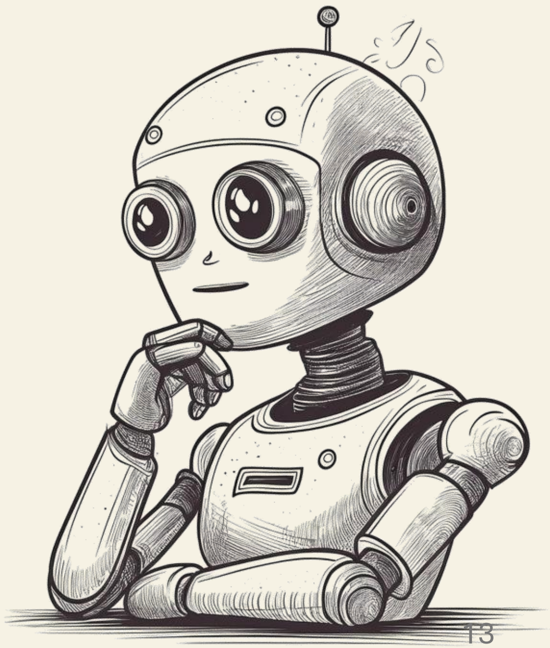
- Simplistic: Can‘t capture fine details.
- Relies on prior understanding.

➔ **Idea:** Learn the form of the differential equation from data!



$$\frac{\mathrm{d}i}{\mathrm{d}t} = f_\theta(i, s)$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = g_\theta(i, s)$$



A Modern, Data-Driven Twist on an Old Story

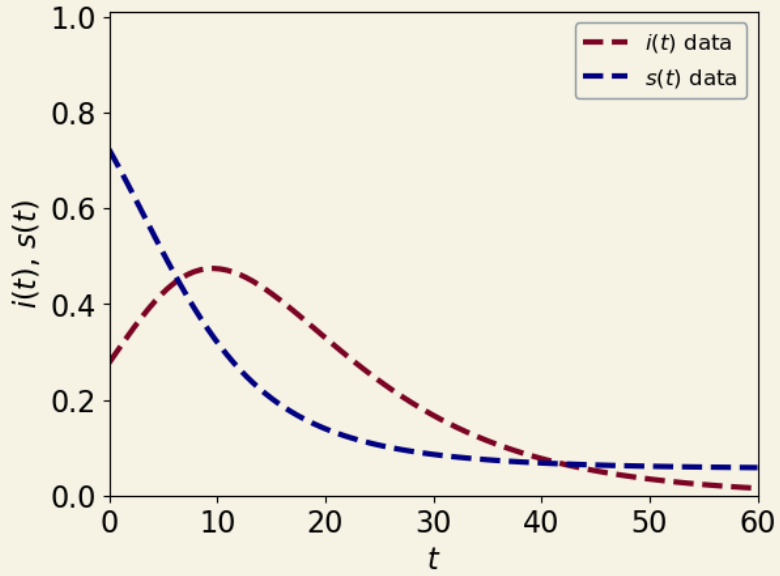
What did we do well?

- Interpretable: All terms have clear meaning.
- Robust: Captures main trends.

What are potential problems?

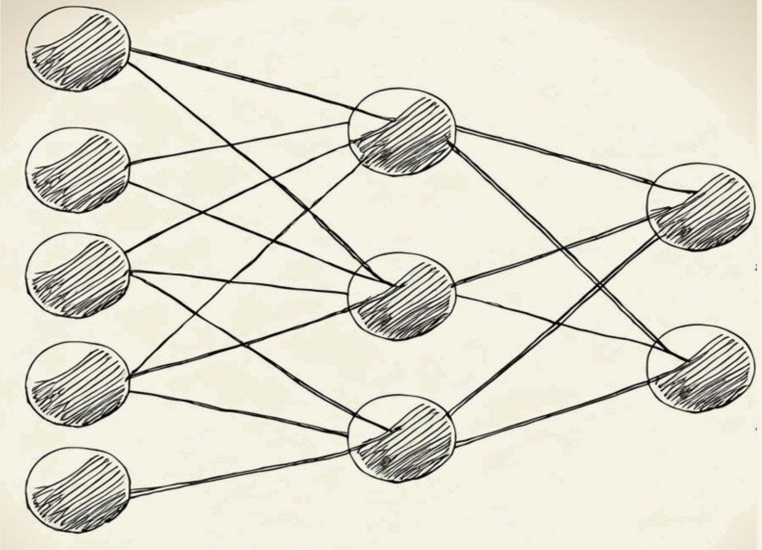
- Simplistic: Can‘t capture fine details.
- Relies on prior understanding.

➔ **Idea:** Learn the form of the differential equation from data!

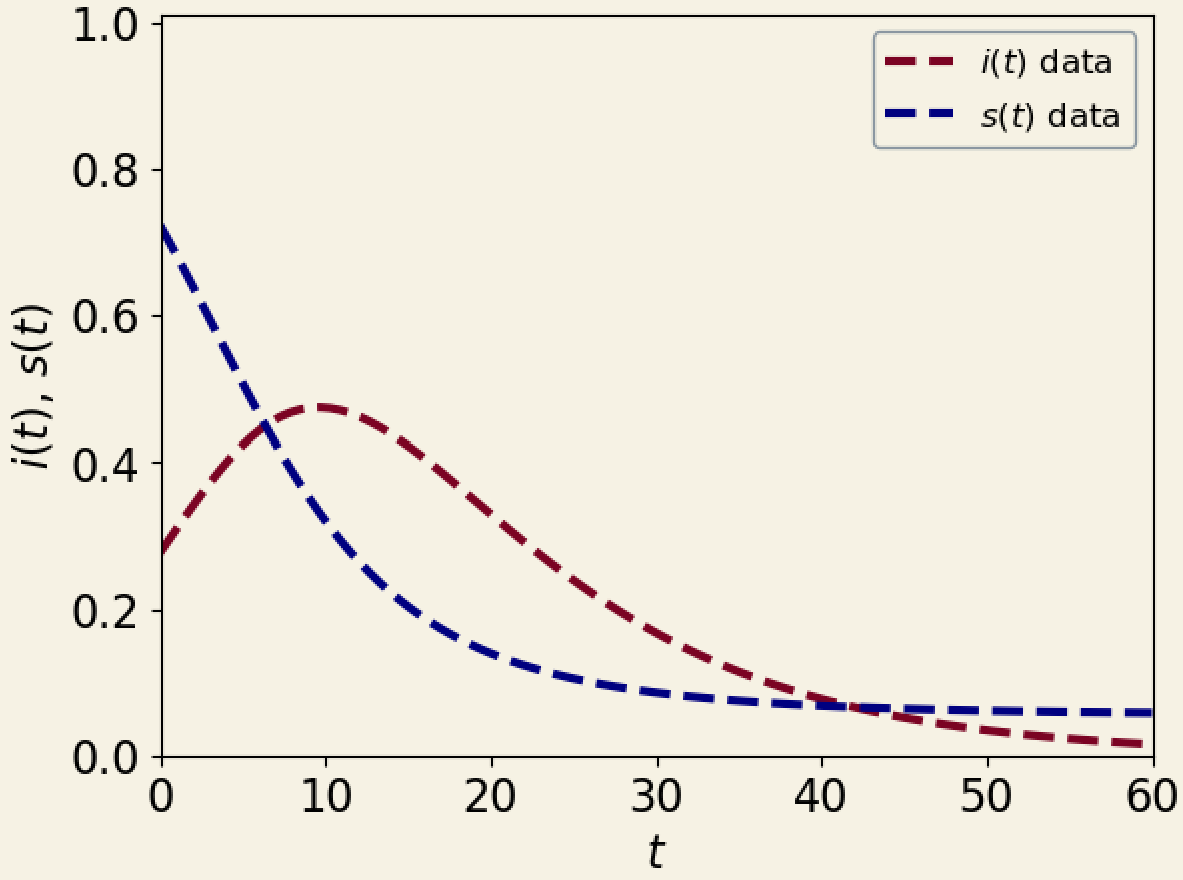


$$\frac{\mathrm{d}i}{\mathrm{d}t} = f_\theta(i, s)$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = g_\theta(i, s)$$



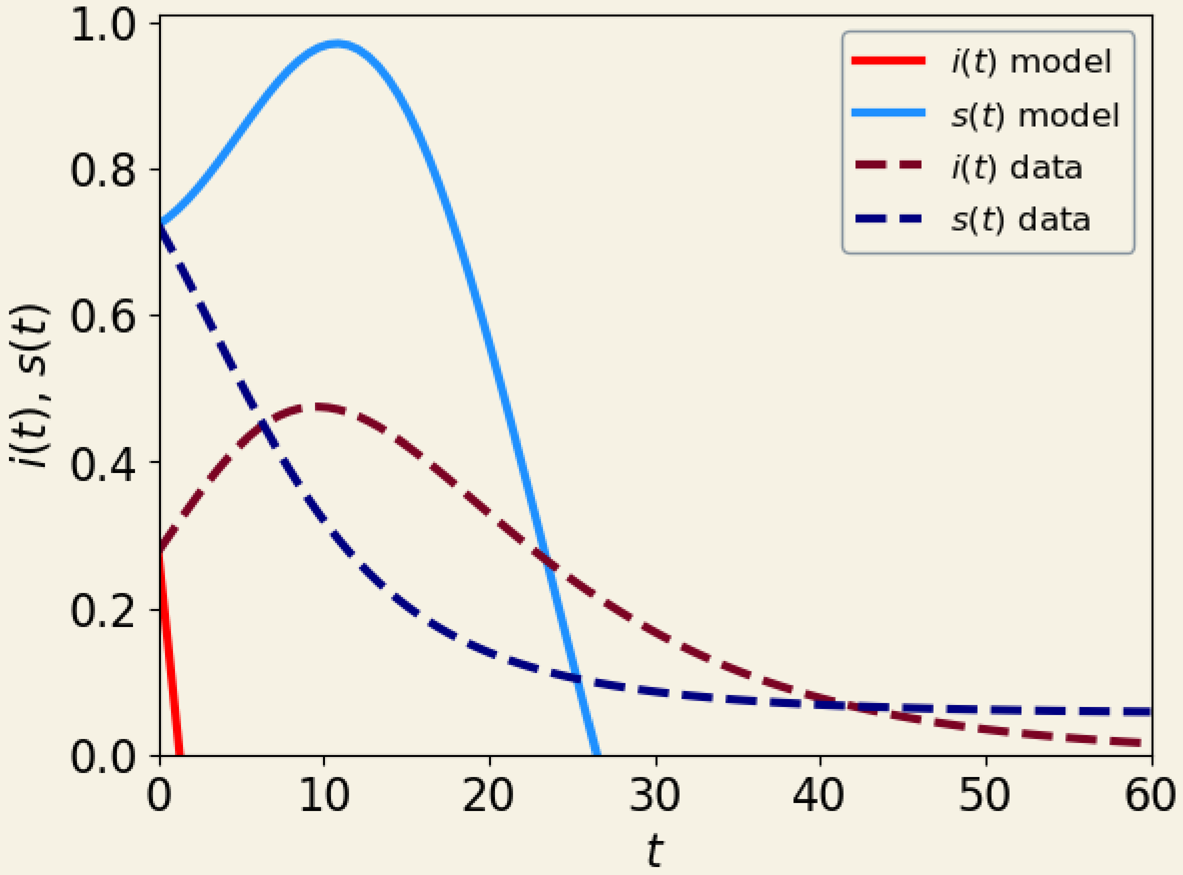
Training Neural ODEs: Broad Strokes



1. Gather Data of the System, usually starting from different initial conditions.

Training Neural ODEs: Broad Strokes

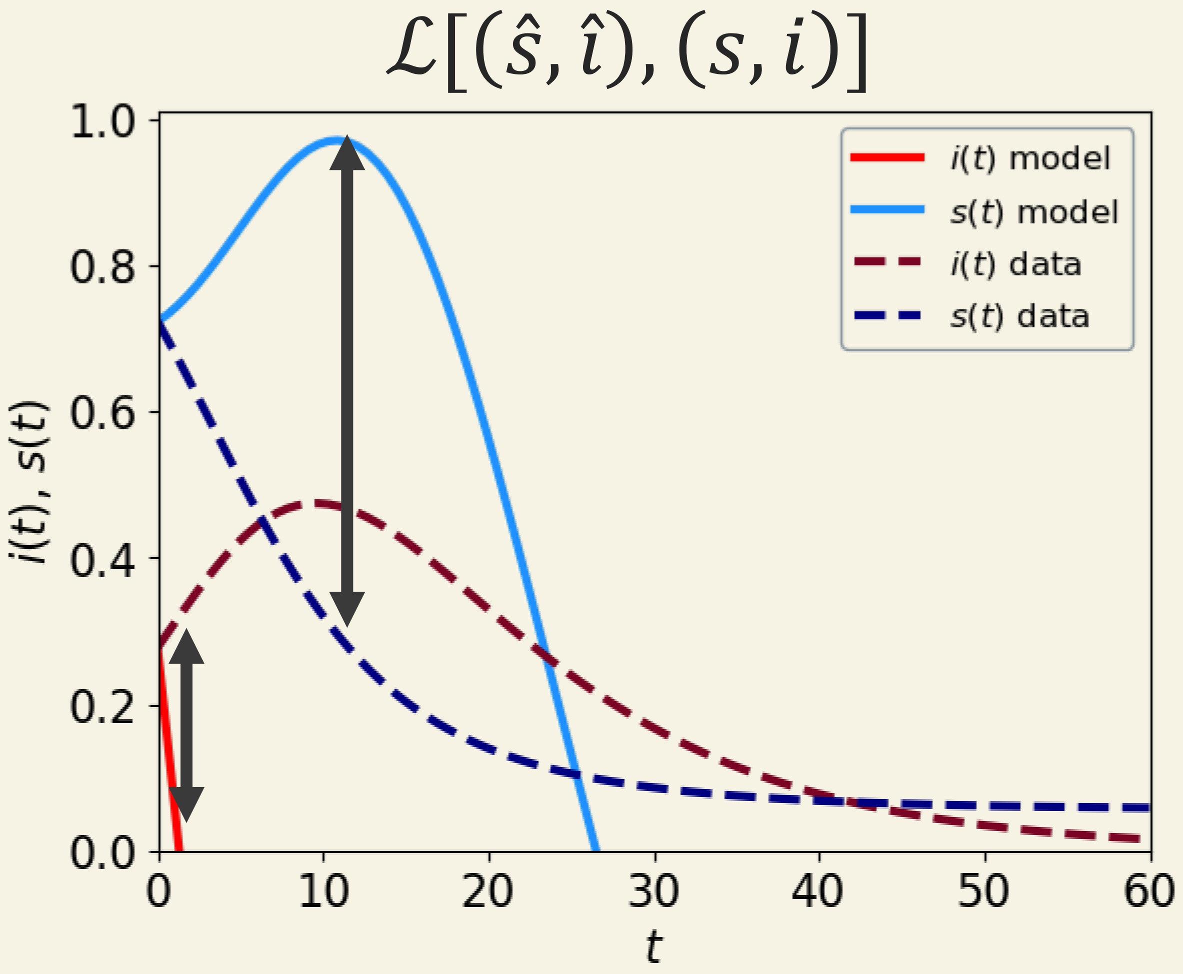
$$(\hat{s}, \hat{\imath})(t) = \texttt{ODEInt}[f_\theta, g_\theta]$$



1. Gather Data of the System, usually starting from different initial conditions.

2. Forward Pass: Solve an ODE with a neural network as right-hand side numerically.

Training Neural ODEs: Broad Strokes

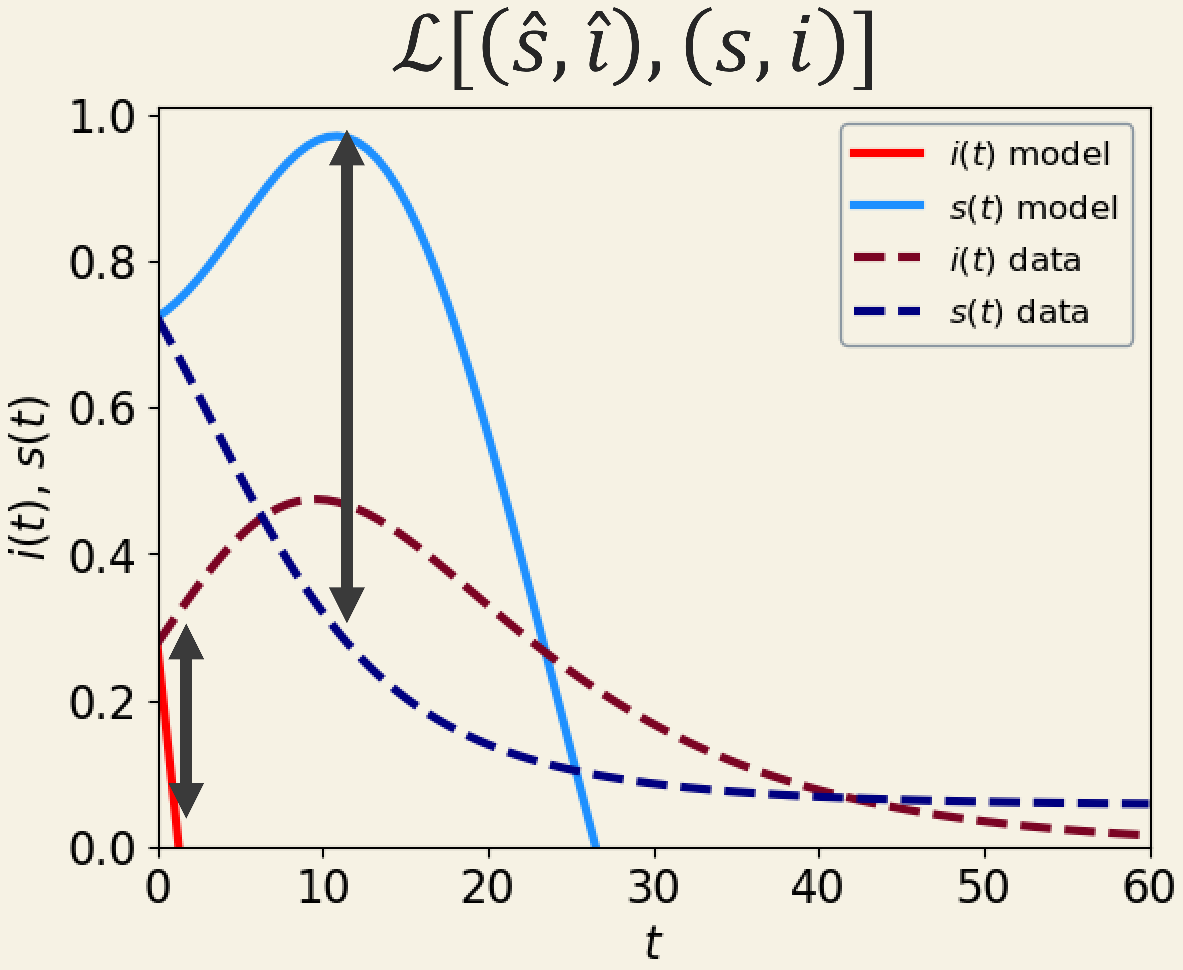


1. Gather Data of the System, usually starting from different initial conditions.

2. Forward Pass: Solve an ODE with a neural network as right-hand side numerically.

3. Compute the loss, e.g., RMSE.

Training Neural ODEs: Broad Strokes



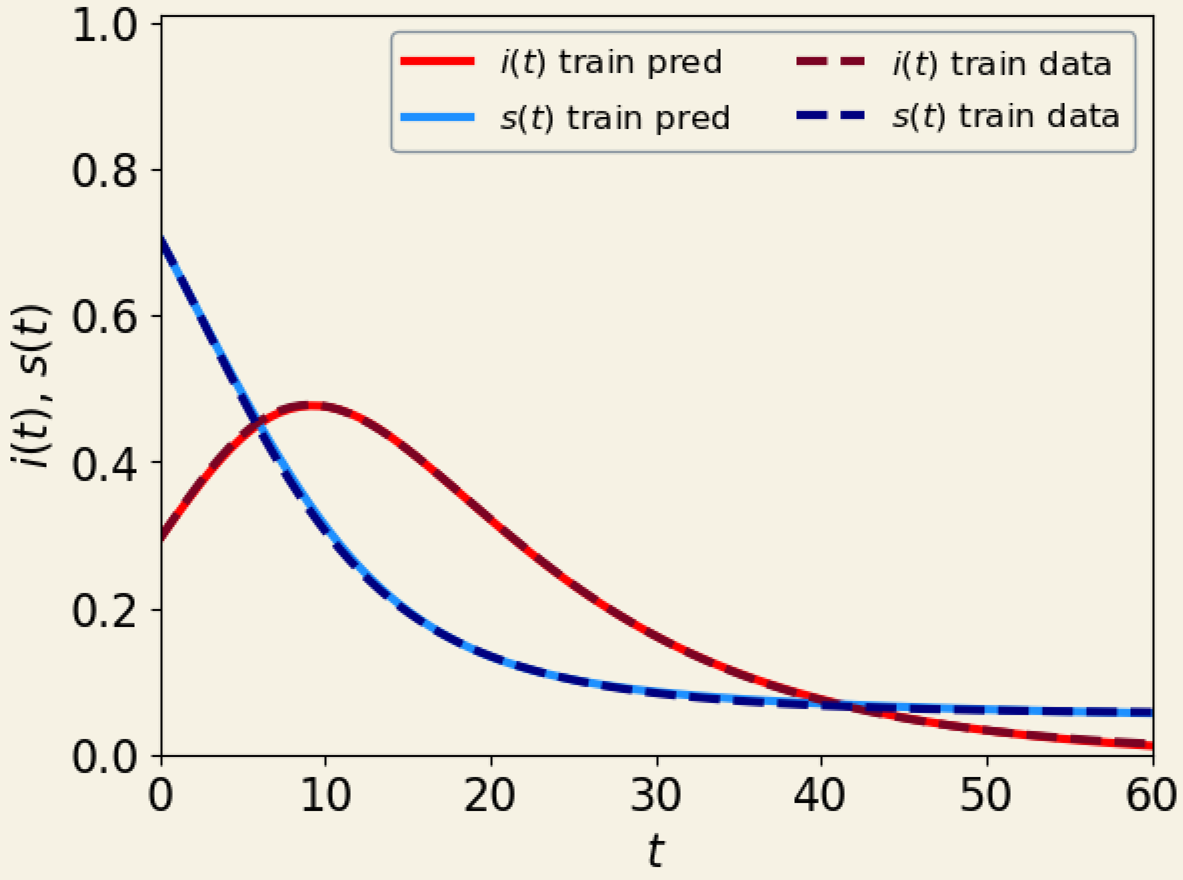
1. Gather Data of the System, usually starting from different initial conditions.

2. Forward Pass: Solve an ODE with a neural network as right-hand side numerically.

3. Compute the loss, e.g., RMSE.

Training Neural ODEs: Broad Strokes

$$\theta \leftarrow \theta + \nabla_\theta \mathcal{L}[(\hat{s}, \hat{\imath}), (s, i)]$$



1. Gather Data of the System, usually starting from different initial conditions.

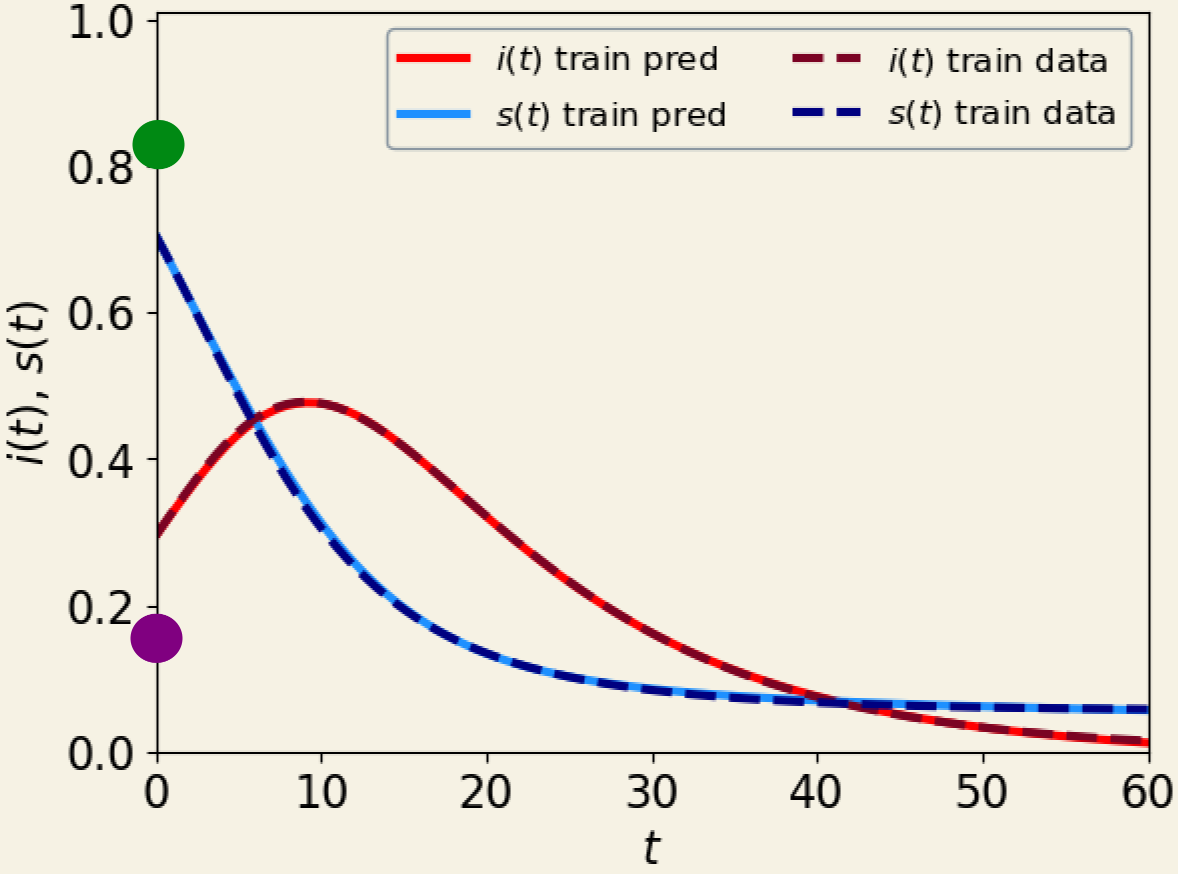
2. Forward Pass: Solve an ODE with a neural network as right-hand side numerically.

3. Compute the loss, e.g., RMSE.

4. Backward Pass: Compute gradients by backpropagation through the ODE solver, and update neural network weights.

Application 1: Prediction

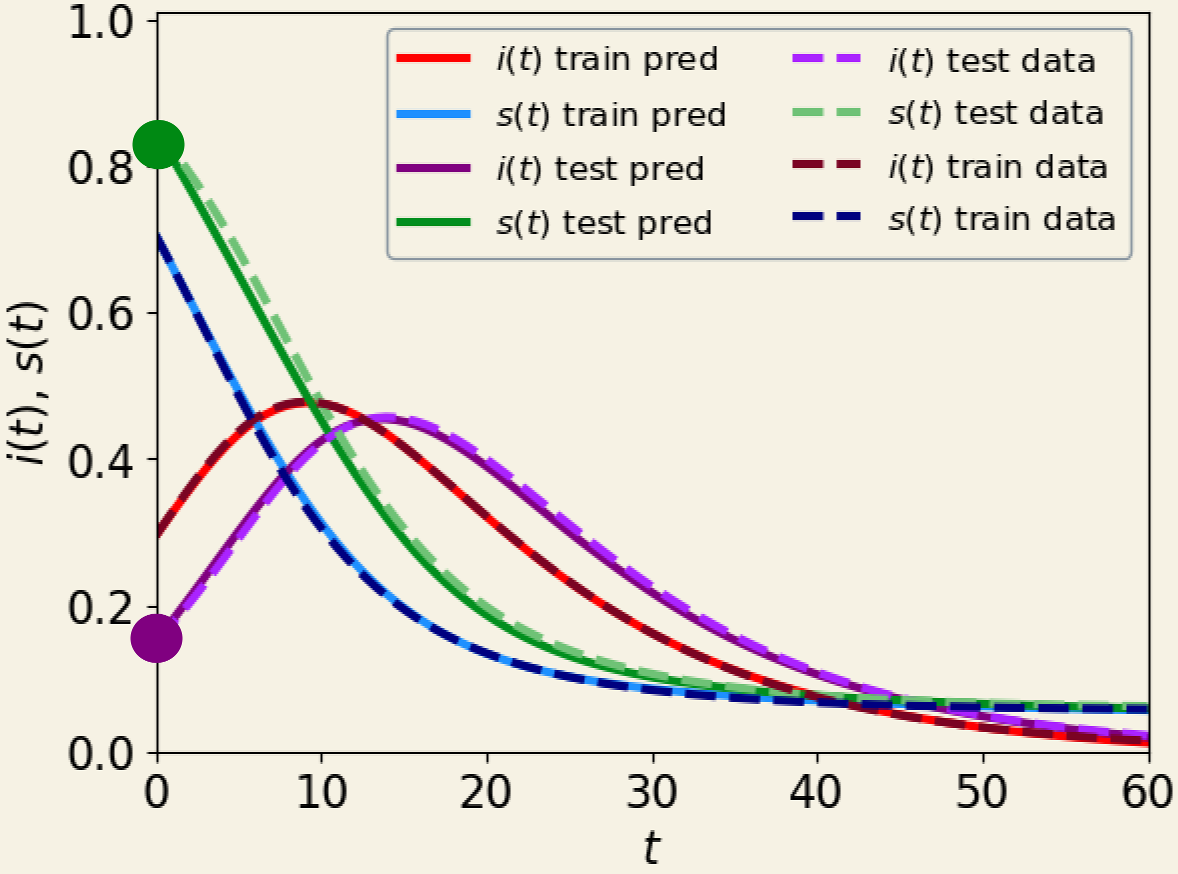
Question: Given the initial condition $i(t=0), s(t=0)$, how will the system evolve?



Application 1: Prediction

Question: Given the initial condition $i(t=0), s(t=0)$, how will the system evolve?

Solution: Feed the initial condition into the neural ODE and use a numerical ODE solver to make a prediction.



Application 2: Combination with ODEs

Question: Can we model a system that we do not fully understand at the relevant scale?

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta i s - \ ??? \ - \mu i(1 - i - s)$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta i s$$

Application 2: Combination with ODEs

Question: Can we model a system that we do not fully understand at the relevant scale?

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta i s - f_\theta(i) - \mu i(1 - i - s)$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta i s$$

Solution: You can combine analytical terms with learned terms to arrive at more accurate descriptions of the system.

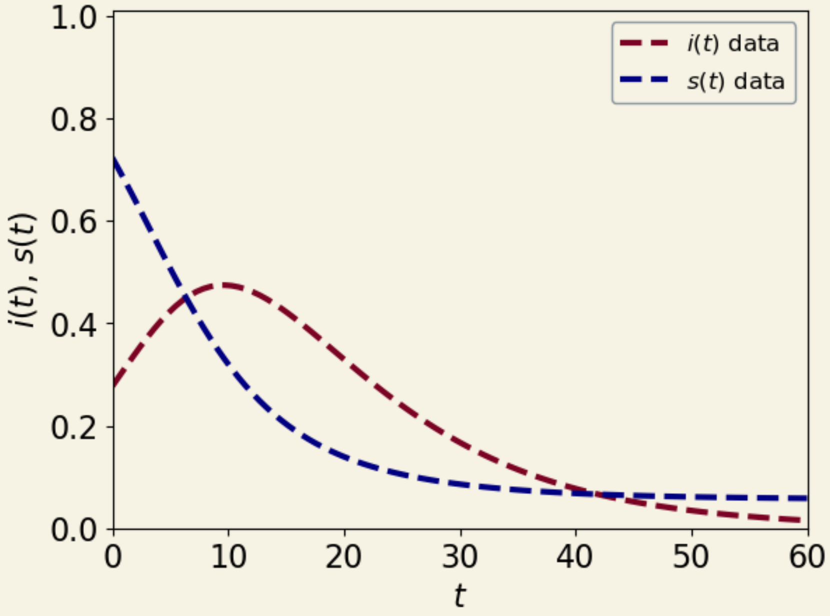
Application 3: Equation Discovery

Question: Can we find the analytical form of an ODE that describes the system?

Application 3: Equation Discovery

Question: Can we find the analytical form of an ODE that describes the system?

We have data.



We want equations.

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta is - \mu i^2 - \mu ir$$

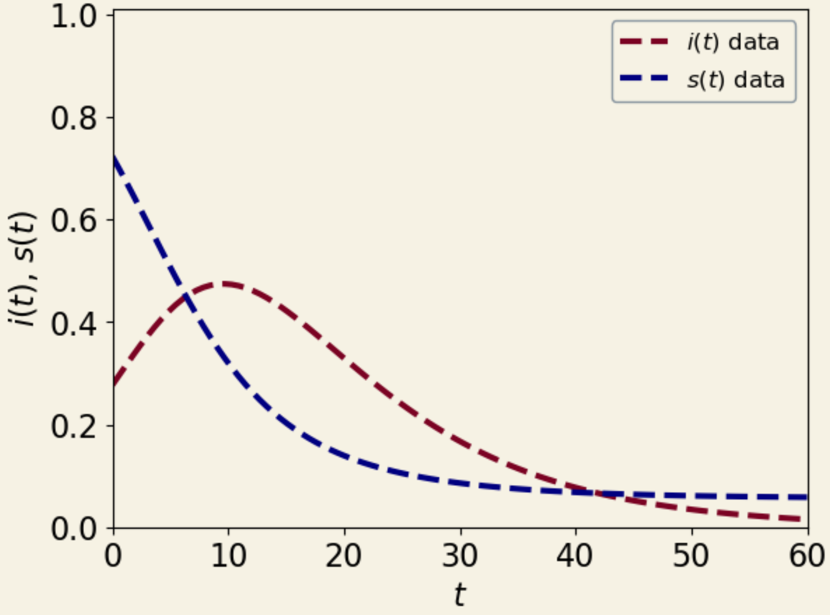
$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta is$$

Application 3: Equation Discovery

Question: Can we find the analytical form of an ODE that describes the system?

Solution: Combine neural ODEs with symbolic regression.

We have data.



We want equations.

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta i s - \mu i^2 - \mu i r$$

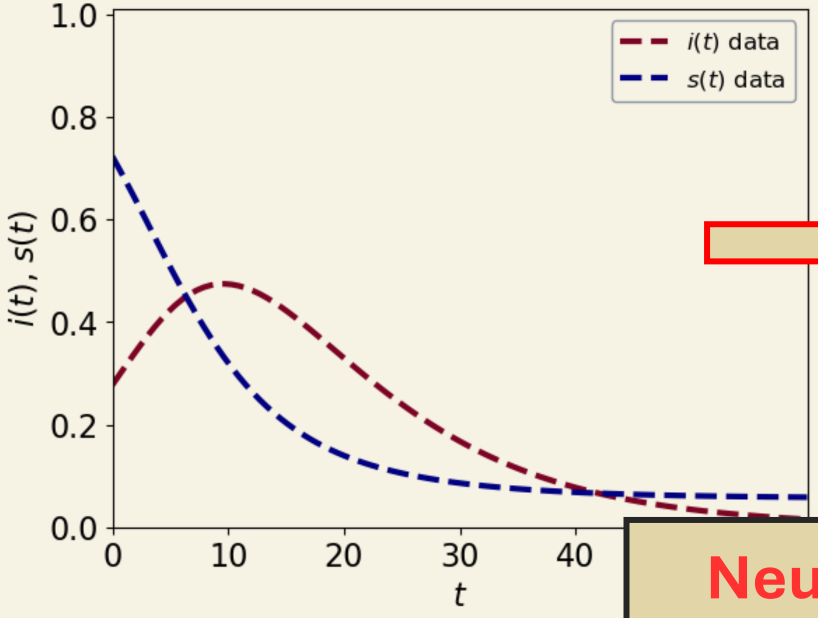
$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta i s$$

Application 3: Equation Discovery

Question: Can we find the analytical form of an ODE that describes the system?

Solution: Combine neural ODEs with symbolic regression.

We have data.



$$\frac{\mathrm{d}i}{\mathrm{d}t} = f_\theta(i, s)$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = g_\theta(i, s)$$

Neural ODE

We want equations.

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta is - \mu i^2 - \mu ir$$

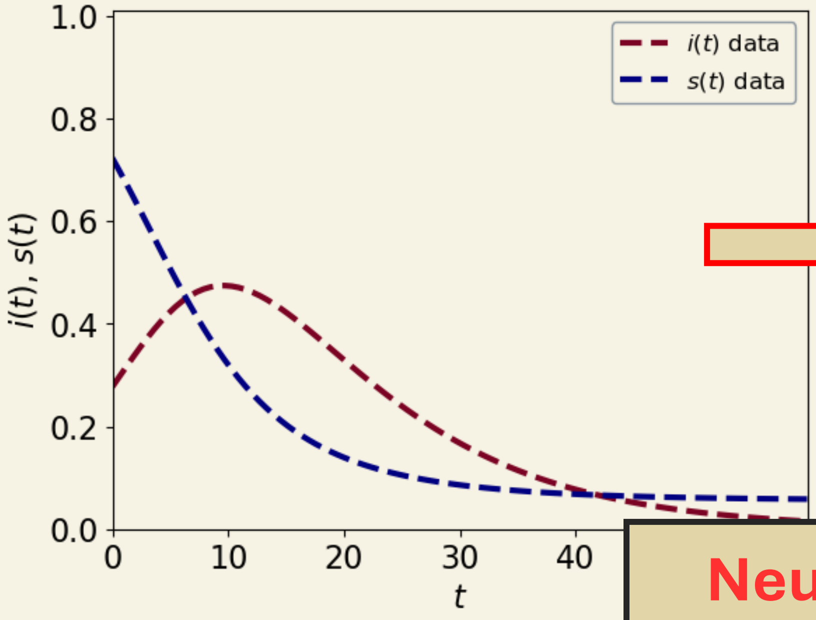
$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta is$$

Application 3: Equation Discovery

Question: Can we find the analytical form of an ODE that describes the system?

Solution: Combine neural ODEs with symbolic regression.

We have data.



$$\frac{\mathrm{d}i}{\mathrm{d}t} = f_\theta(i,s)$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = g_\theta(i,s)$$

Neural ODE

We want equations.

$$\frac{\mathrm{d}i}{\mathrm{d}t} = \beta is - \mu i^2 - \mu ir$$

$$\frac{\mathrm{d}s}{\mathrm{d}t} = -\beta is$$

Symbolic Regression

Want to know more?

Chen et al. (2018) is popularized neural ODEs in the modern sense.

Want to know more?

Chen et al. (2018) is popularized neural ODEs in the modern sense.

Patrick Kidger's PhD Thesis *On Neural Differential Equations* and this talk of his are great starting points to learn about neural ODEs.

Want to know more?

Chen et al. (2018) is popularized neural ODEs in the modern sense.

Patrick Kidger's PhD Thesis *On Neural Differential Equations* and this talk of his are great starting points to learn about neural ODEs.

I personally use jax, diffrax and equinox for neural ODEs but a pytorch option called torchdiffeq or torchdyn also exists.

Want to know more?

Chen et al. (2018) is popularized neural ODEs in the modern sense.

Patrick Kidger's PhD Thesis *On Neural Differential Equations* and this talk of his are great starting points to learn about neural ODEs.

I personally use jax, diffrax and equinox for neural ODEs but a pytorch option called torchdiffeq or torchdyn also exists.

Shameless self-plug: Soon, you'll find a paper on the topic on my google scholar. 😏

Hands On Neural ODEs

Intro to jax

[UvA DL Notebooks]

[jax tutorial]

Neural ODEs

[diffrax neural ODE tutorial]

[Uva DL Notebooks]

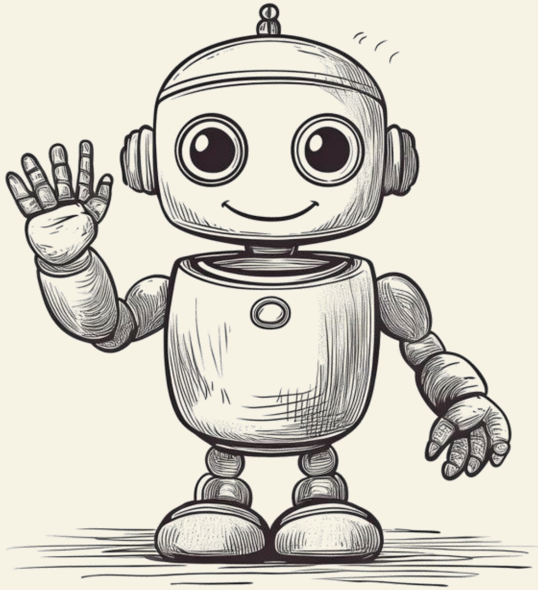
or klick

[here]

Other Tutorials

[diffrax tutorials]

Thank you.



Copyright

Laber M., Klein B., Eliassi-Rad T. *When do neural ordinary differential equations generalize on complex networks?* arXiv:2602.08980 (2026)