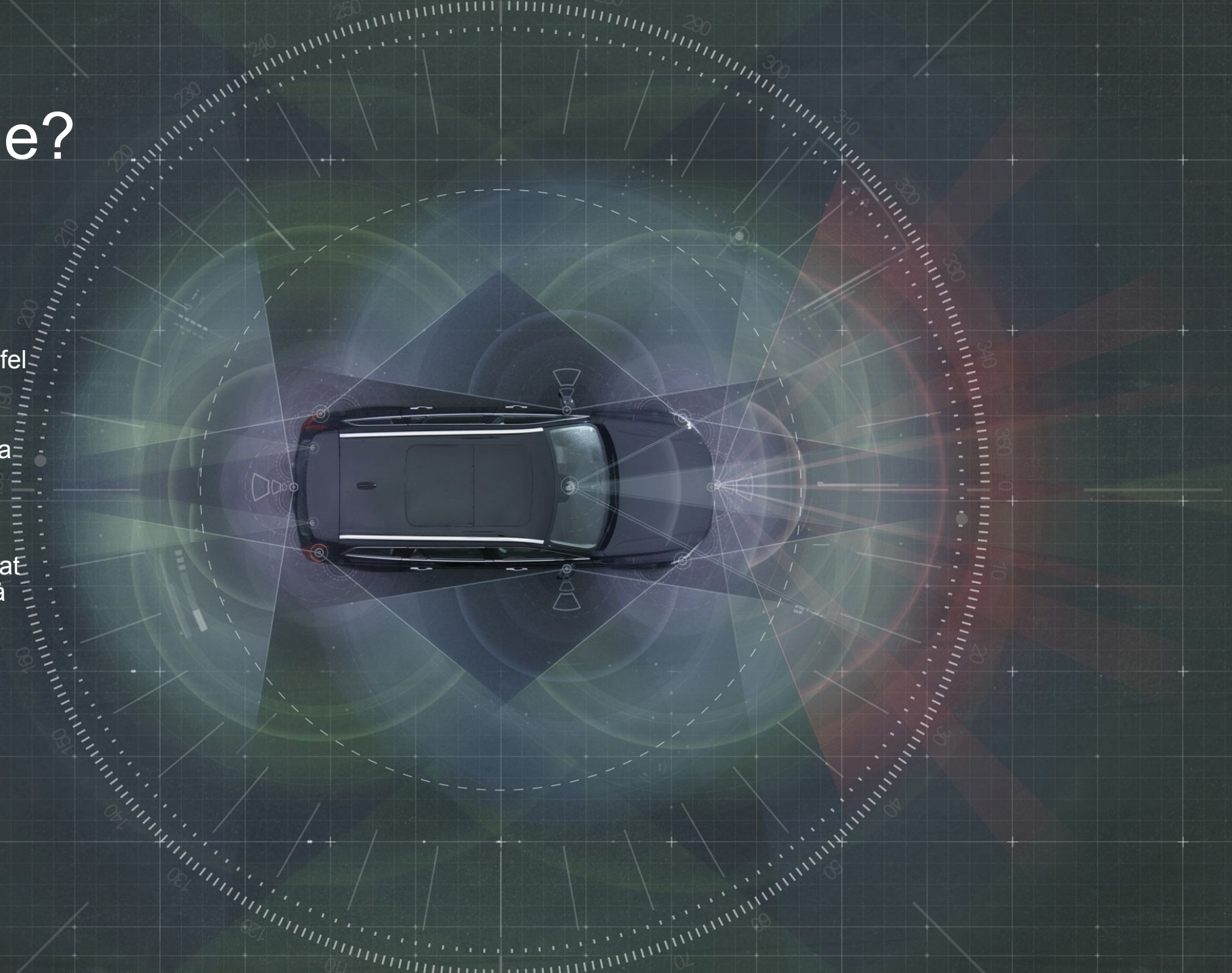
# Självkörande bilar: Maskininlärning och matematik

Edvin Listo Zec
edvin.listo-zec@zenuity.com

ZENUITY

Make it real.

# Varför självkörande?



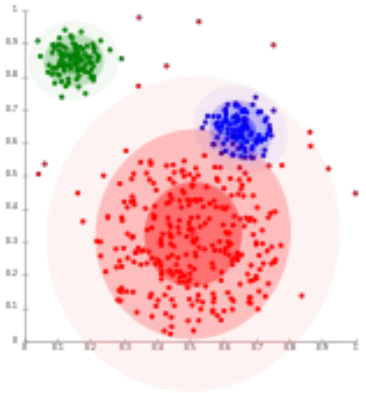
- **Säkrare**: de flesta olyckor i trafiken sker pga mänskliga fel
- Bilen kan användas av andra när du inte använder den
- **Spara tid:** Fokusera på annat på väg till jobbet (läsa, se på Netflix, etc.)
- Kollektivtrafik
- Lastbilar och transport

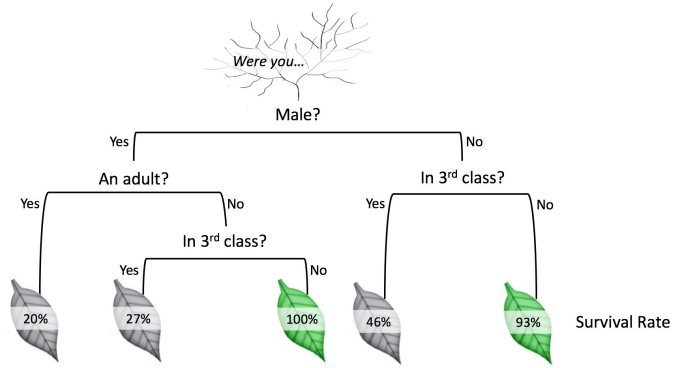
I mitt jobb fokuserar jag mycket på ***maskininlärning***: Algoritmer som med hjälp av data “lär sig själva” genom att hitta mönster och samband.

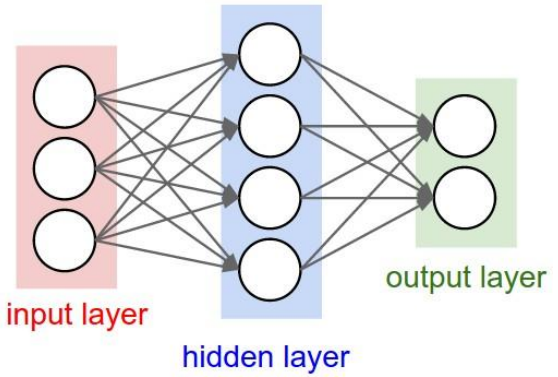
Maskininlärning kombinerar **statistik**, **linjär algebra** och **algoritmer**.

Exempel på maskininlärningsalgoritmer:

- Decision Trees
- Neural network
- Hidden Markov models
- Clustering
- Många fler...






ZENUITY

# Exempel på där maskininlärning används:

- Netflix, Amazon, Spotify: rekommenderar produkter utifrån vad du har tittat på tidigare
- Google Deepmind: Utvecklat AlphaGo, ett program som genom att spela mot sig själv slagit världens bästa spelare i Go. Satsar nu på att utveckla liknande algoritmer för spelet Starcraft 2.
- Inom bildbehandling: Klassificera objekt i en bild. Te.x. om man ser en fotgängare eller en cyklist, eller inom sjukvård klassificera typ av cancer.
- Röstigenkänning: Siri, Google Home etc.
- Prediktion: Förutspå framtiden (finans, självkörande bilar)

- **Zenuity** är ett nytt företag: ett joint venture mellan **Volvo Cars** och **Autoliv**
  - Advanced Driver Assistance Systems (ADAS)
  - Highly Autonomous Drive (HAD)

- På Zenuity på Lindholmen är vi ca 400 anställda fördelade i många olikar grupper som tillsammans utvecklar mjukvara för självkörande bilar.

- Exempel på grupper:

  Collision avoidance, Object fusion, Deep learning, Ground truth och många fler.

- Mitt team: **Data Analysis Team**


ZENUITY

# Det jag jobbar med just nu:

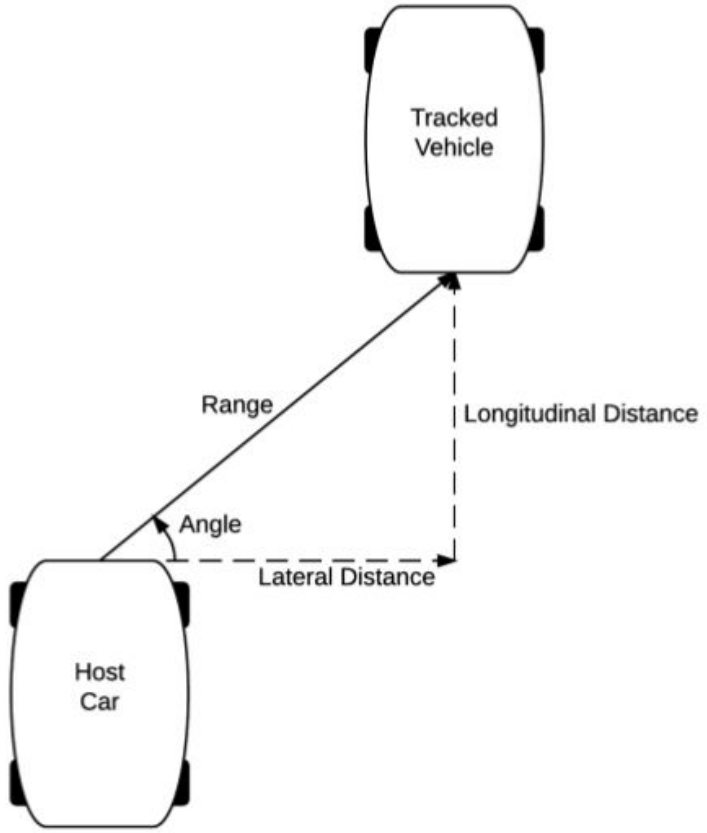
Att med hjälp av maskininlärning försöka förstå och **modellera felet i sensorerna** som används i självkörande bilar.

Min modell lär sig den underliggande **sannolikhetsfördelningen**

hos felet. Det innebär att man kan *generera artificiellt fel.*

Varför artificiellt fel?

- Kostar tid och pengar att åka med bilar och samla in data
- Går att använda artificiella datan i simuleringar för tester




ZENUITY

Datan som jag jobbar med är stor:

En bils insamlade data kan bestå av 50 kolumner av intressanta variabler. Exempelvis:

- Position
- Hastighet
- Acceleration

...med ca 2 miljoner rader som representerar varje tidssteg man mäter.

Matrisen av data: $A \in \mathbb{R}^{2 \cdot 10^6 \times 50}$

Detta innebär att förutom matematik så behöver man **effektiva algoritmer** som kan hantera så stora matriser**.**

# Hidden Markov Model

## Konkret inledande exempel:

Betrakta Agneta och Bengt som bor i olika städer. Bengt är endast intresserad av tre saker:

- Promenera (walk)
- Shoppa (shop)
- Städa hemma (clean)

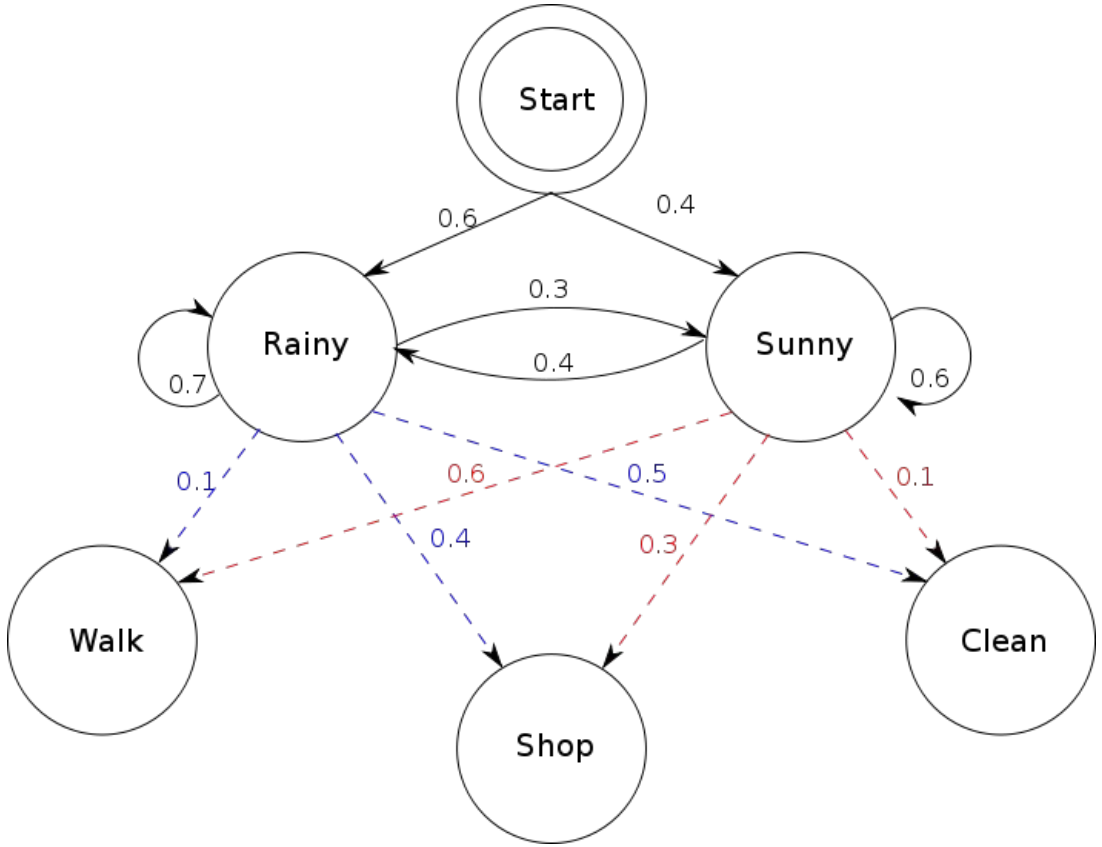
Vad han gör bestäms endast utav vädret.

Agneta som bor i en annan stad har ingen aning om vad vädret är (rainy vs sunny), detta är “hidden”.

Hon vet dock generella trender: tex sannolikheten att det blir en regnig dag efter en solig dag.

Agneta får information från Bengt om vad han gör en given dag (observationer).

Detta system är per definition en Hidden Markov model.



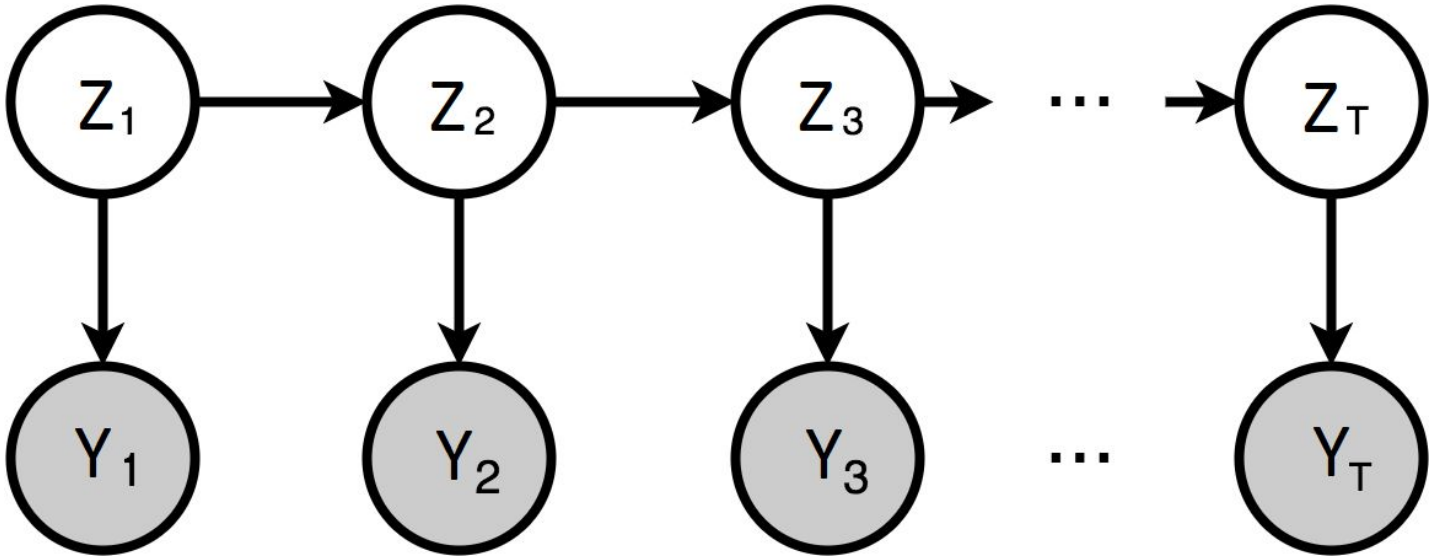
# Hidden Markov Models

Populär algoritm som används mycket när man analyserar tidsserier.

$Y_t$ Observation vid tid *t*
$Z_t$ Hidden State vid tid *t*

$$p(Z_{t+1} = n|z_1, z_2, \ldots, z_t) = p(Z_t = n|z_t)$$

$$p(y_{t+1}|y_1, y_2, \ldots, y_{t-1}, z_1, z_2, \ldots, z_{t-1}) = p(y_t|z_t)$$



# Hidden Markov Model

Definieras av tre parametrar:

1. Initiala sannolikheten att börja i olika tillstånd.

$$\boldsymbol{\pi} = [\pi_1, \ldots, \pi_n]$$

2. Övergångsmatrisen
   (sannolikheten att gå från ett tillstånd till ett annat)

$$\boldsymbol{A} \in \mathbb{R}^{n \times n}$$

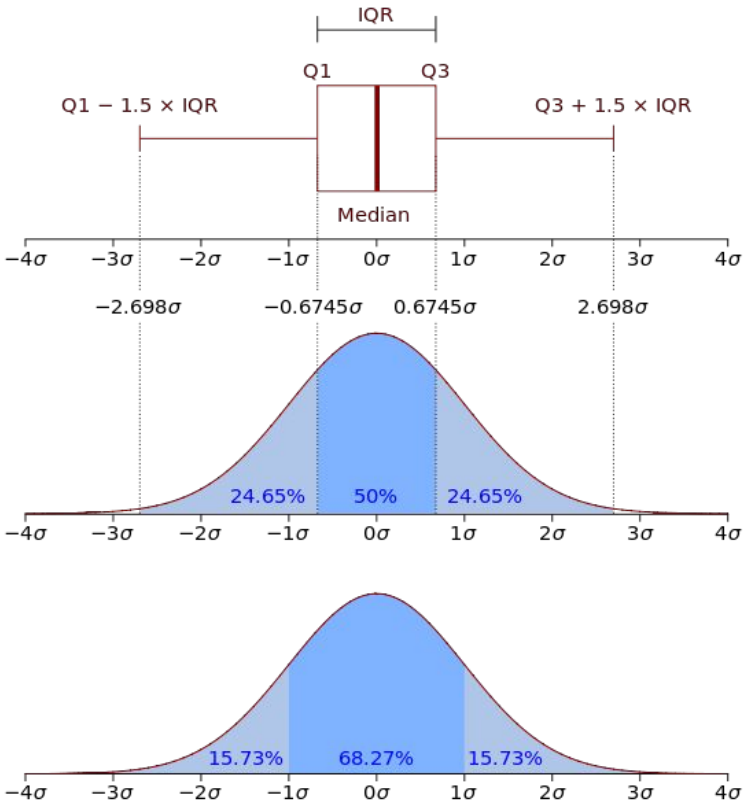
3. Sannolikheten för observationerna i varje tillstånd

$$p_j(y)$$

# Hidden Markov Models

Algoritmen **lär sig iterativt med hjälp av data** att hitta den underliggande **sannolikhetsdistributionen**.

När man har lärt sig sannolikhetsdistributionen kan man ta stickprov från den och på så sätt **generera artificiell data**.




ZENUITY

# Autoregressive Input/Output Hidden Markov Model

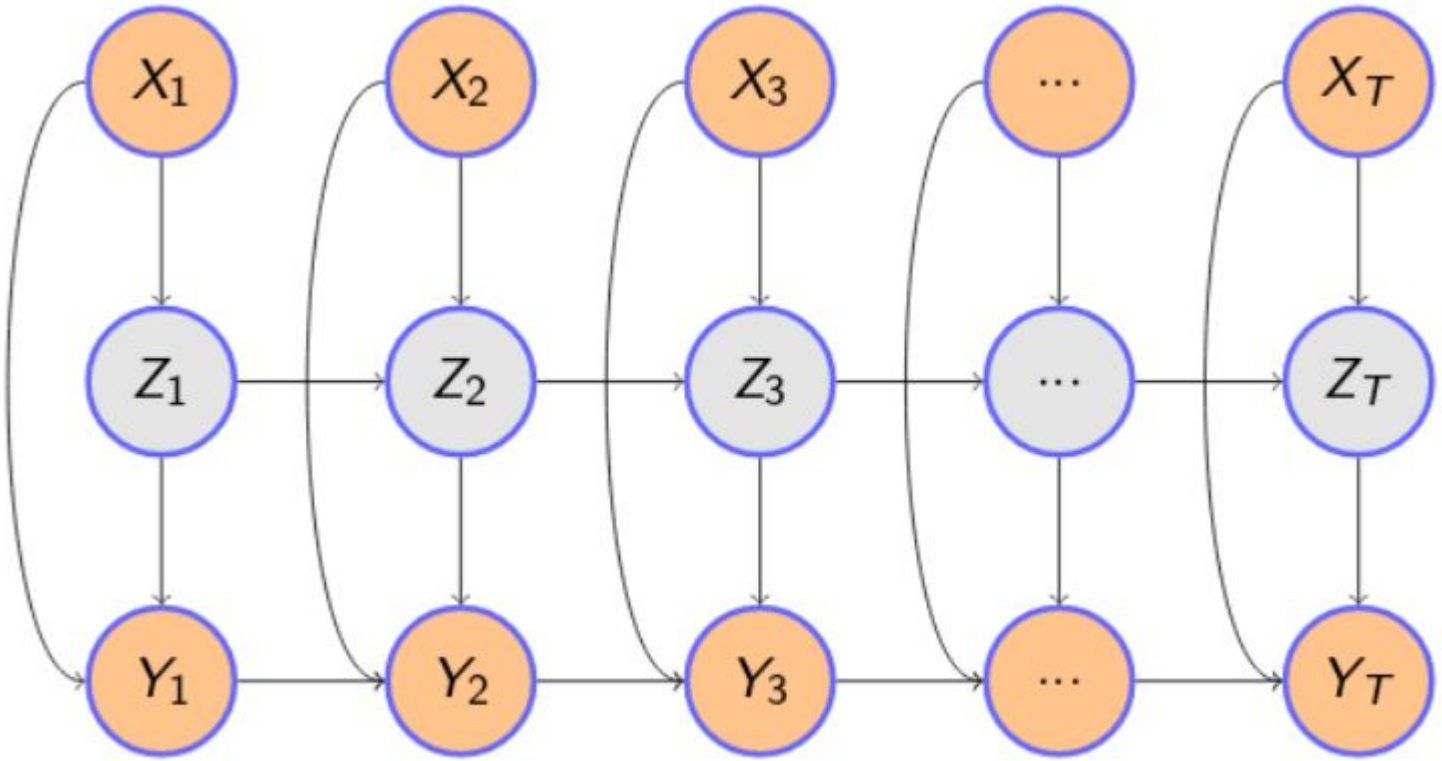
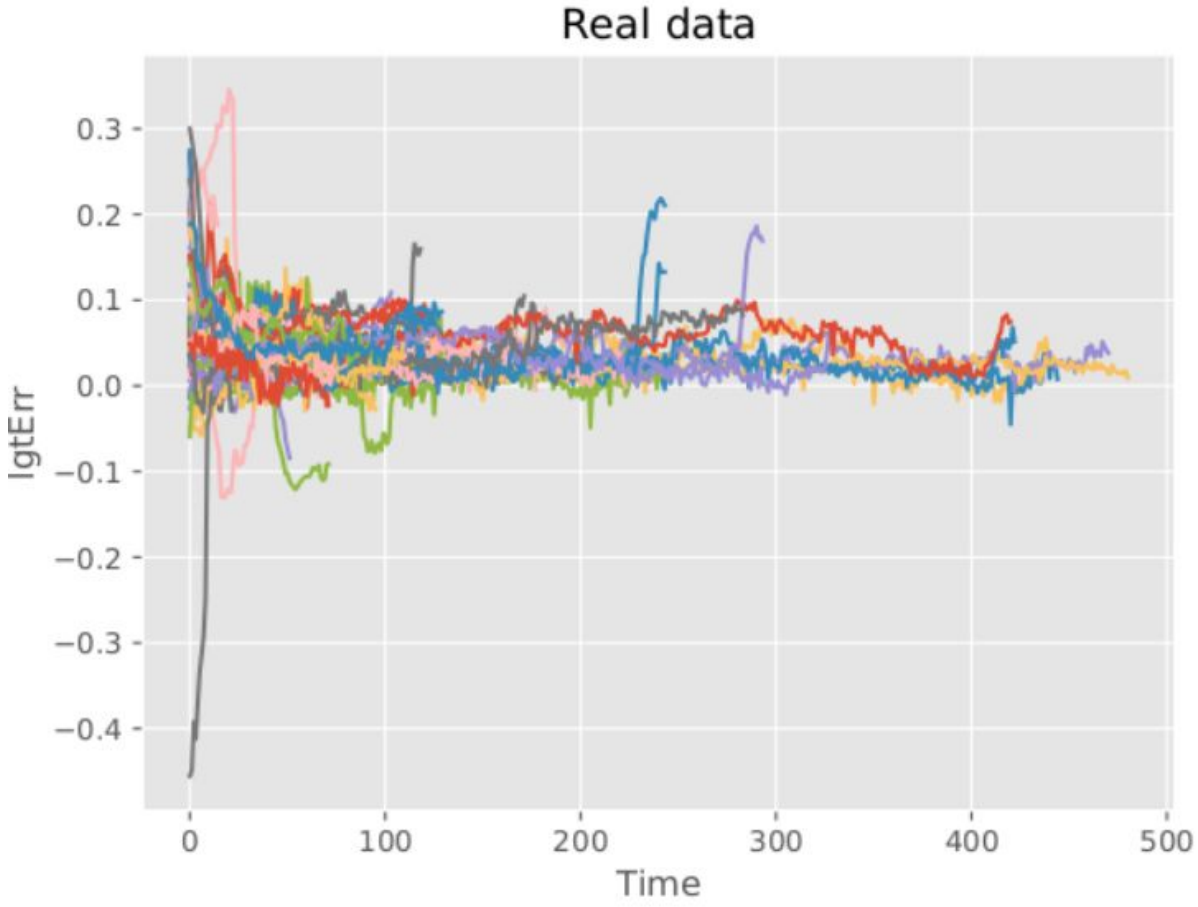


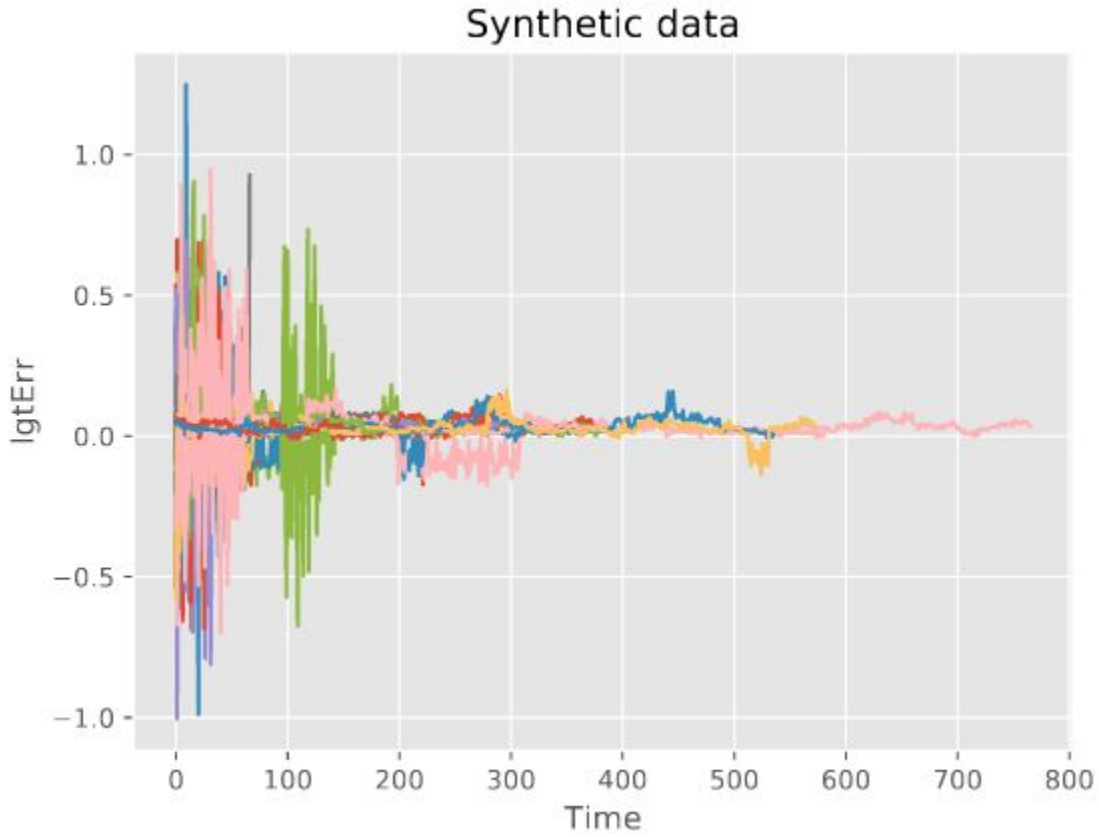

Figure: Graphical representation of an AIOHMM

X: Input
Z: Hidden state
Y: Output

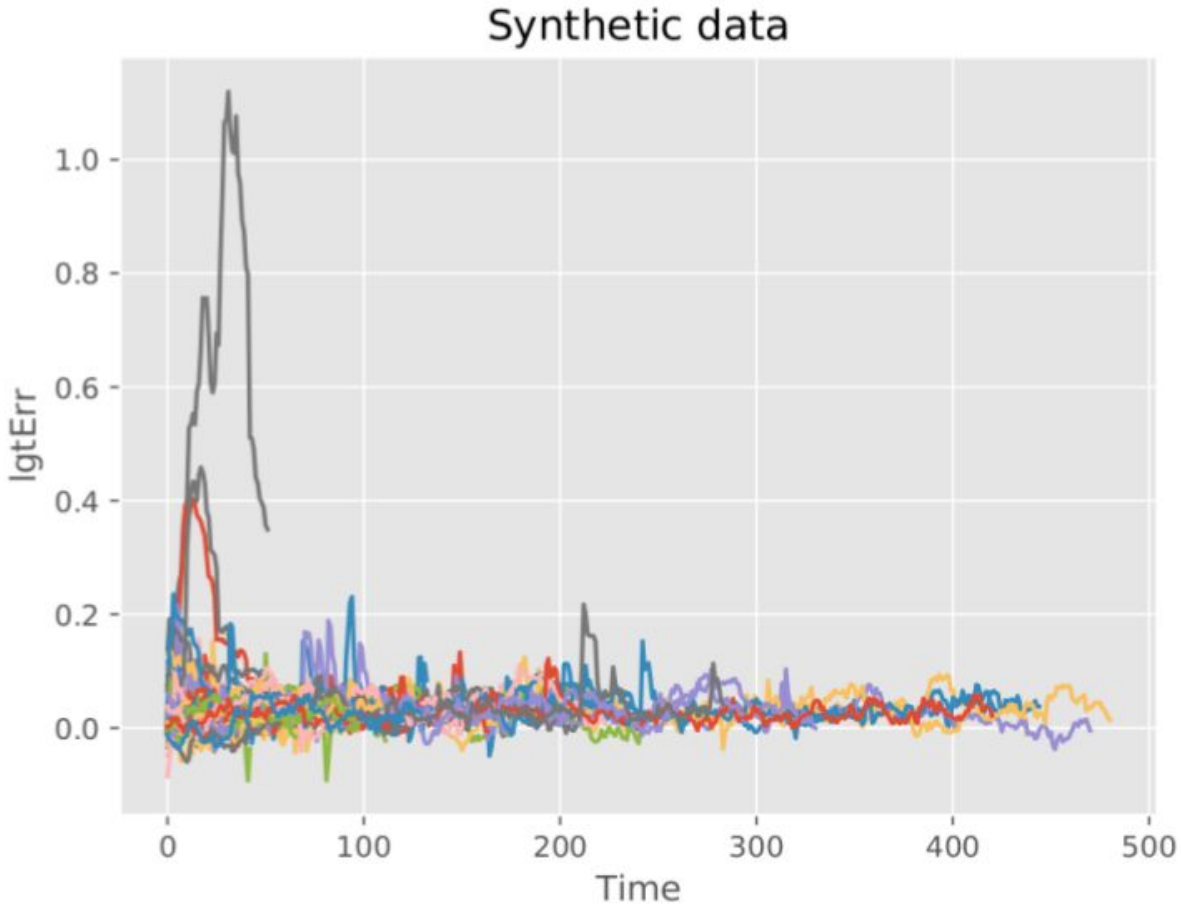
# Datan jag vill modellera så nära som möjligt



# En väldigt dålig Hidden Markov Model i tidigt stadium

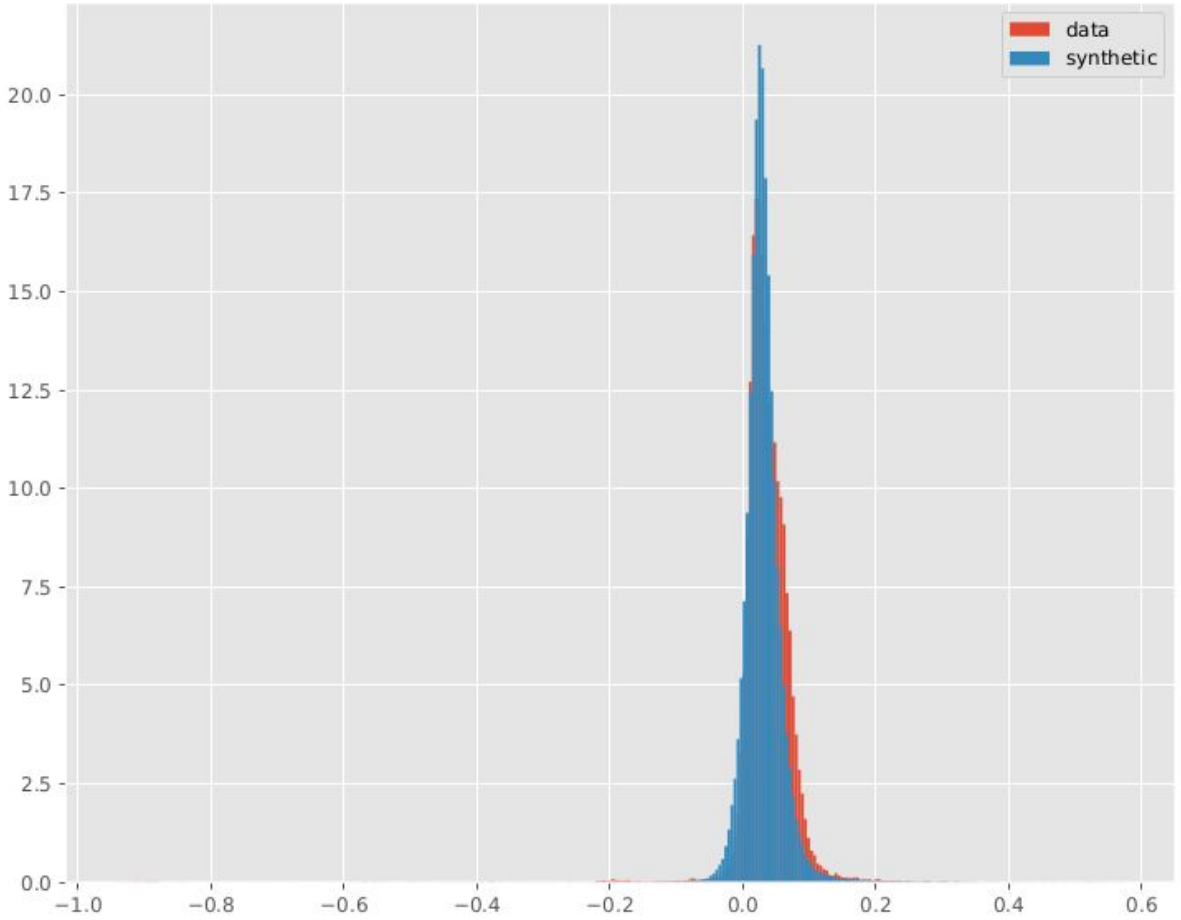


# Den förbättrade modellen




ZENUITY

# Histogram över syntetisk data och riktig data




ZENUITY

# Frågor?

ZENUITY

Make it real.