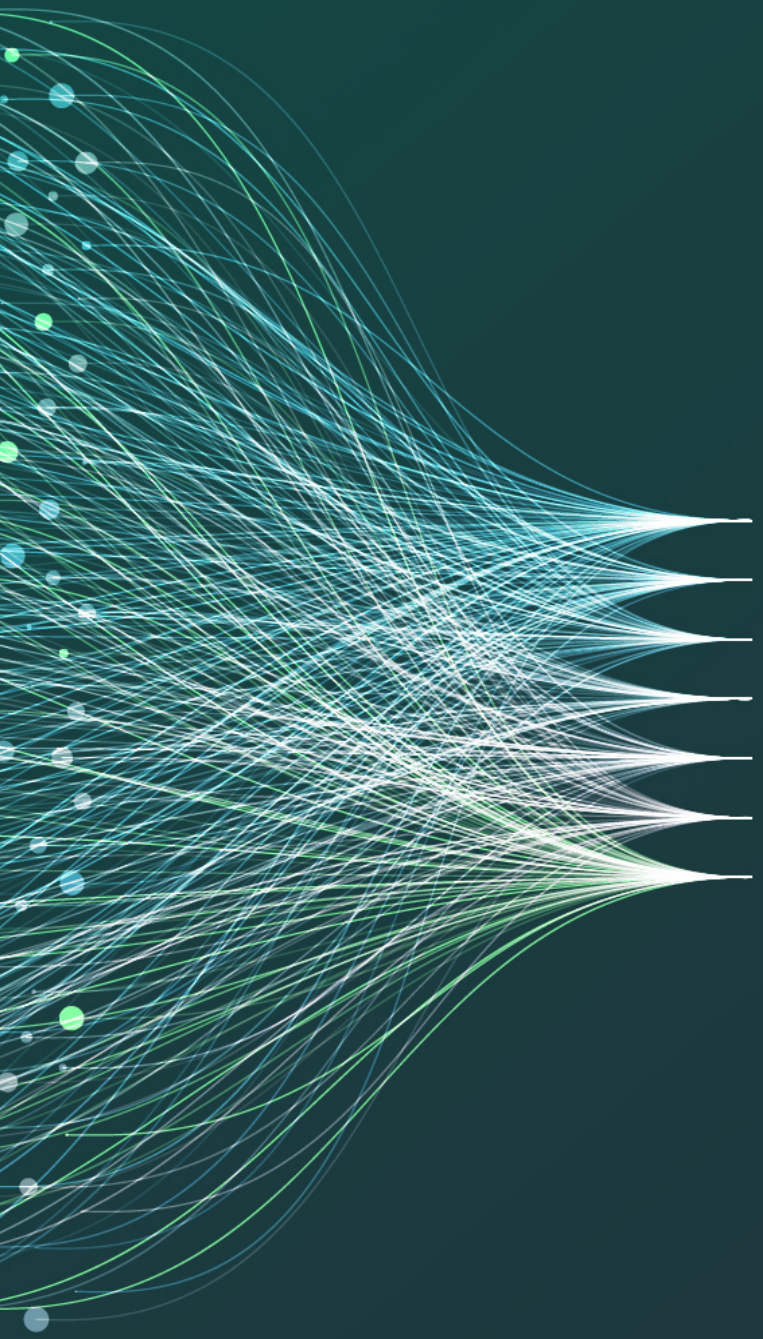
Automating Machine Learning to Solve Real World Problems

Sari Andoni, PhD

sparkcognition™



sparkcognition™

What is Automated Machine Learning? (AutoML)

A brief history of Learning

Psychology:

Learning is a process that depends on experience and leads to long-term changes in behavior potential.

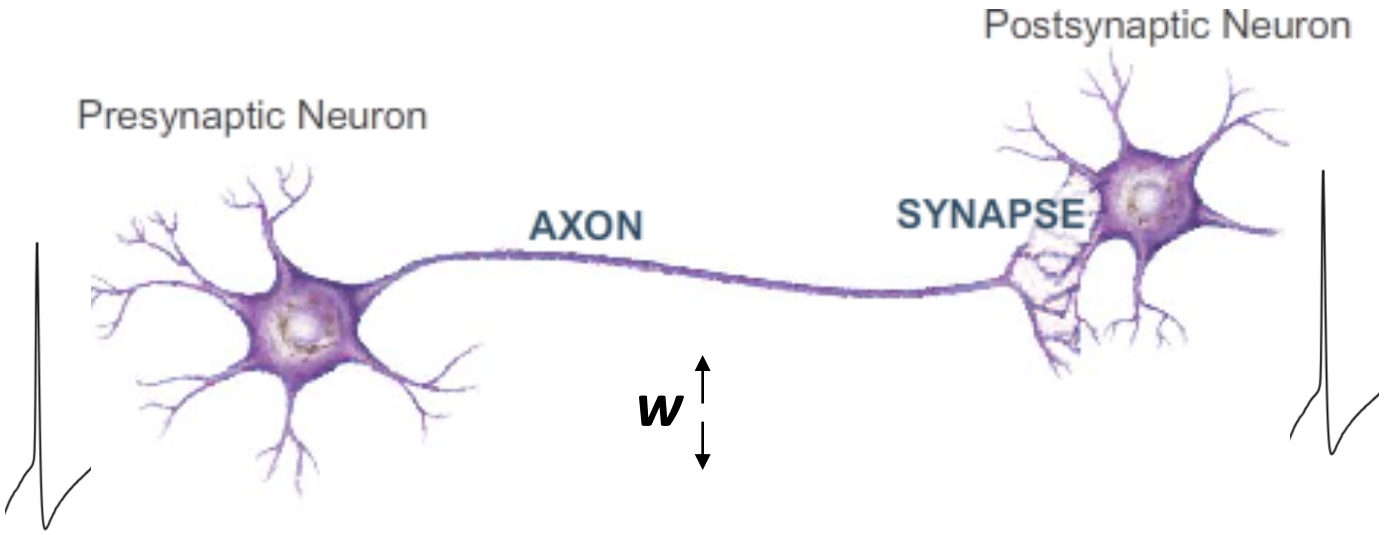
Learning Types:

- Classical conditioning [Pavlov 1890s]
- Operant conditioning (reward & punishment) [Skinner 1938]
- Observational Learning (Imitation) [Bandura 1977]

A brief history of Learning

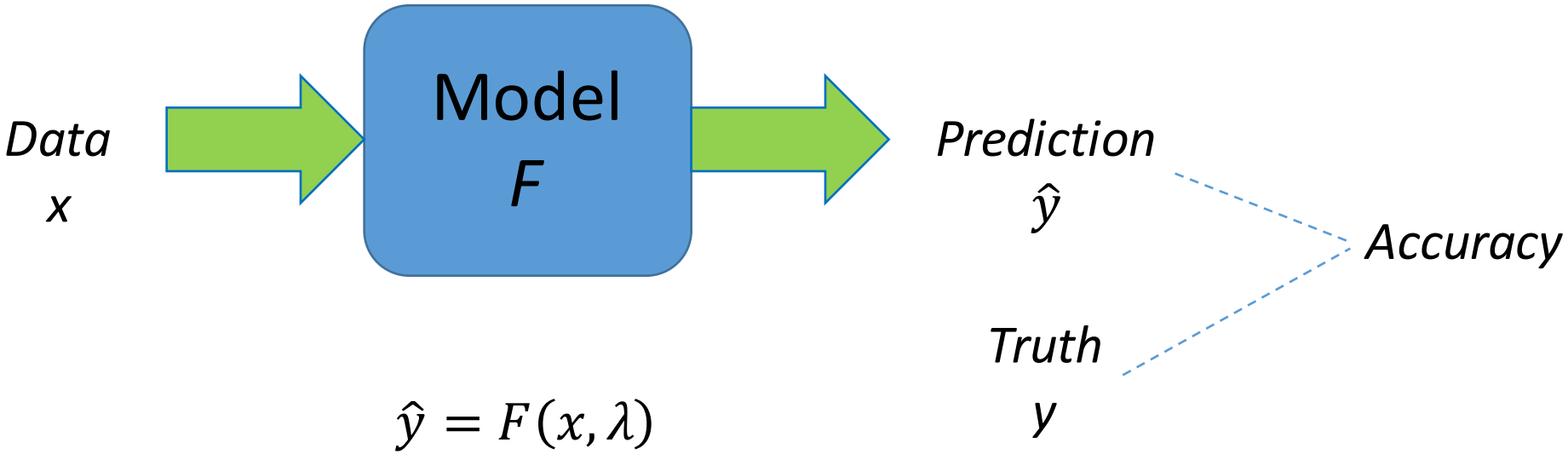
Neuroscience:

- Hebbian Learning: Neurons that fire together wire together [Hebb 1949]
- Long-term synaptic potentiation and depression [Kendal 1976]
- Attractor networks, e.g. Hopfield Network



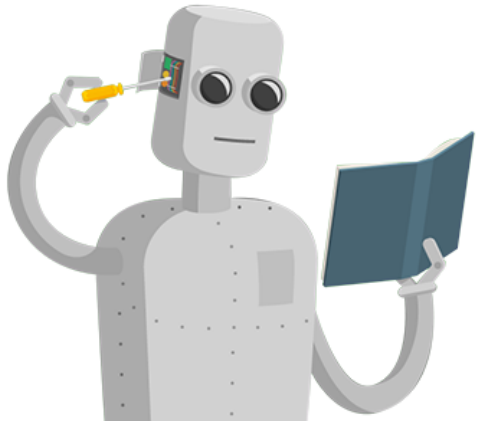
A brief history of Learning

Machine Learning

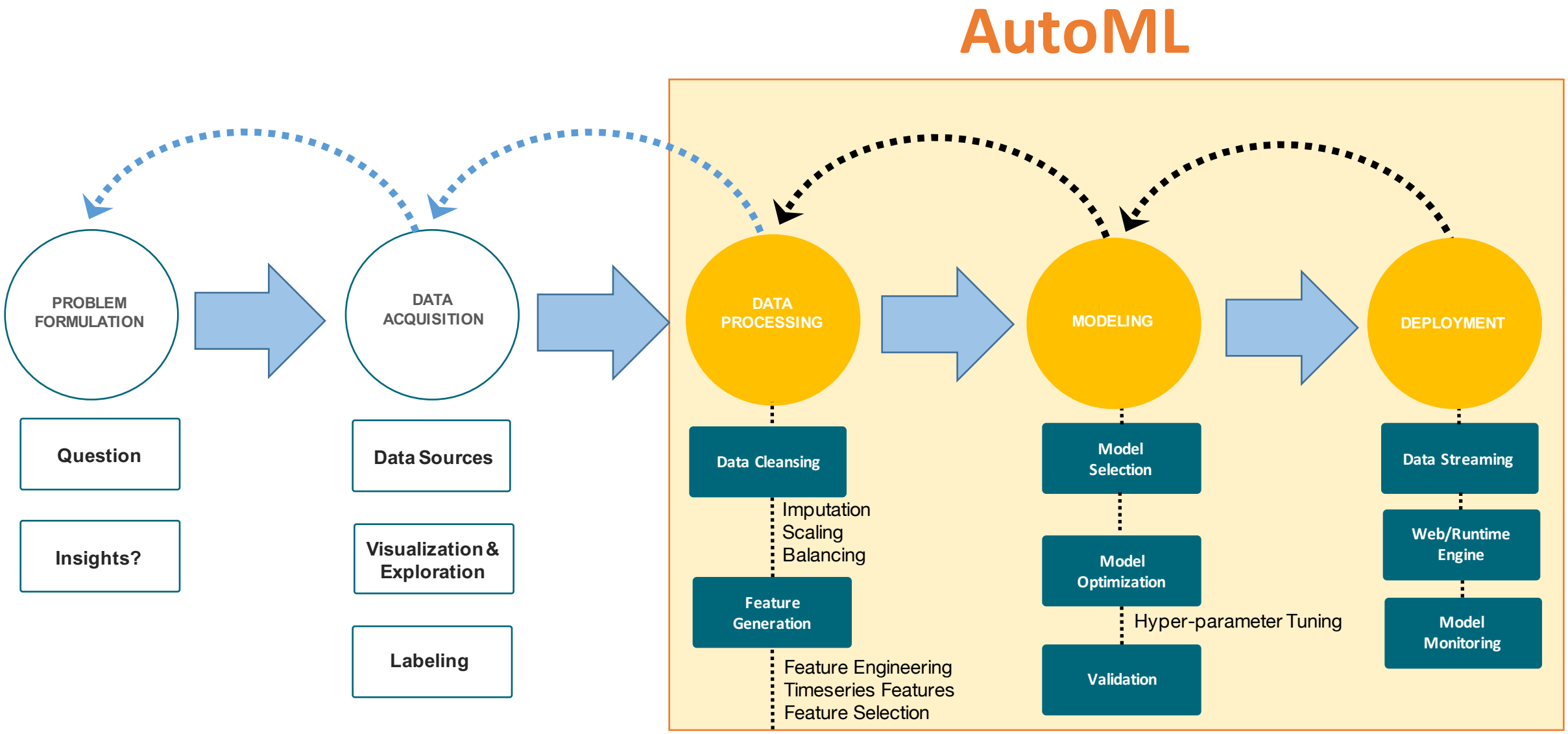


$$\hat{y} = F(x, \lambda)$$

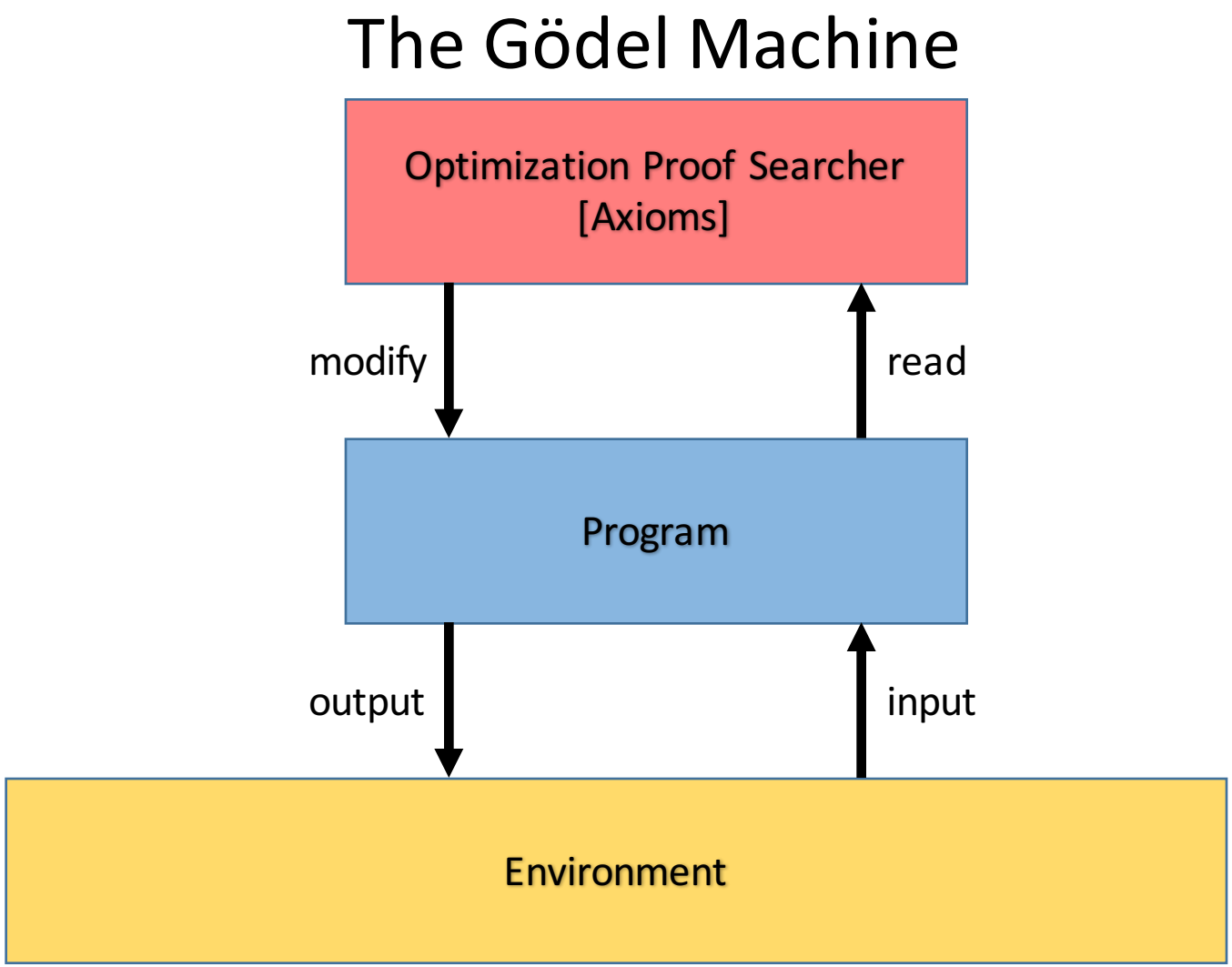
Find F, λ that $\min Obj(y, \hat{y})$

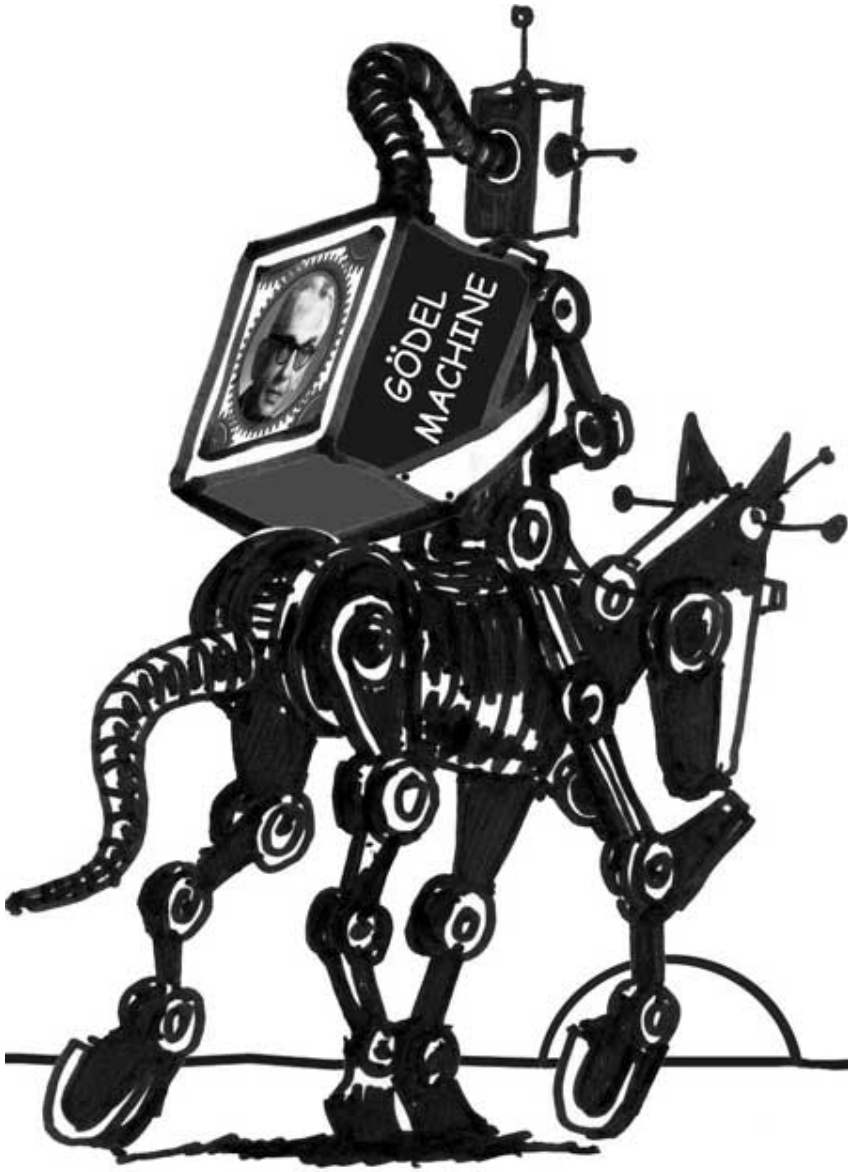


The Machine Learning (DS) Process



Can the Learning *Process* be Machine only?





Jürgen Schmidhuber 2003

Why do we Automate Machine Learning?

Data, Data Everywhere!

How do we scale the Data Science process?

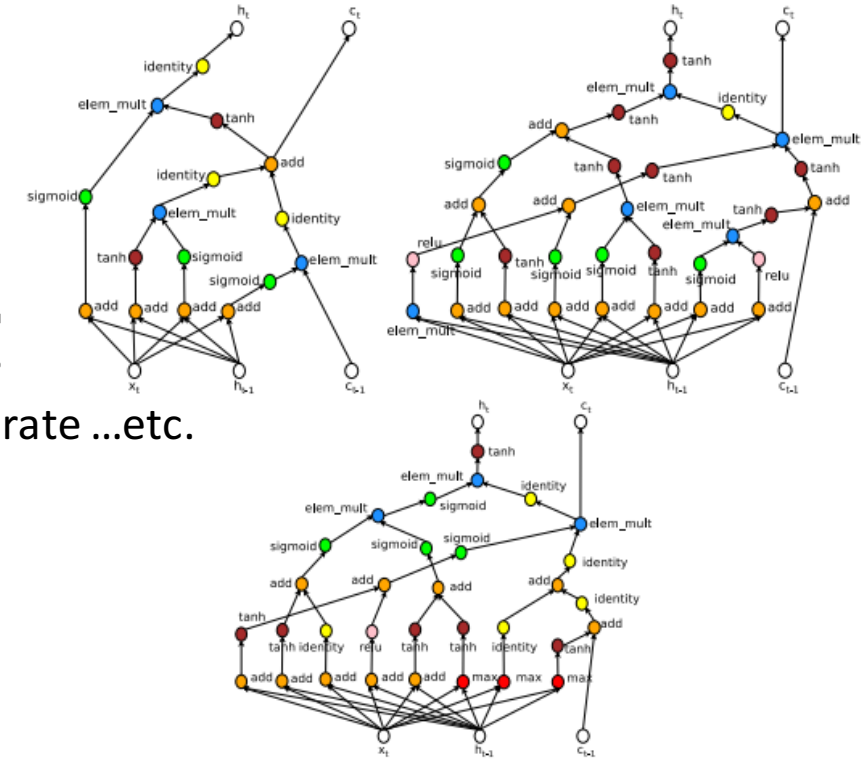
1. Accessibility
Maximize scale of data science teams by making model building accessible to everyone

2. Automation
Automate accurate machine learning models that solve a diverse range of problems

3. Self Maintenance
Quickly adapt to changing conditions over time with automated model retraining

How can we automate Deep Learning?

Step 1: Architecture Search



Step 2: Hyper-parameter Tuning

batch size, dropout, SGD algorithm, learning rate …etc.

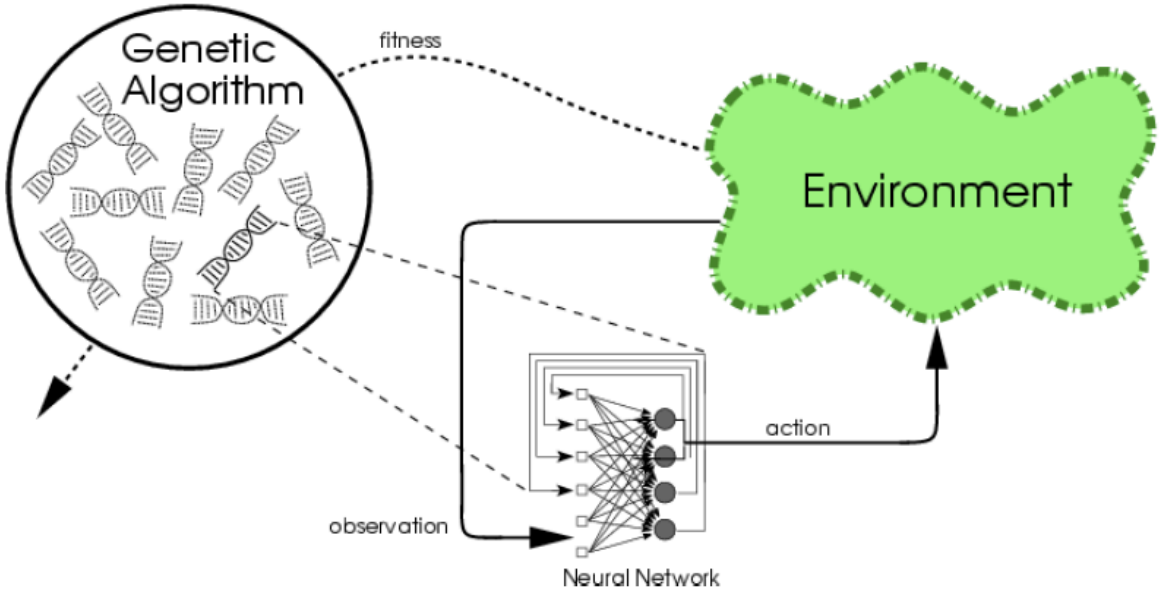
Step 3: Training

backpropagation

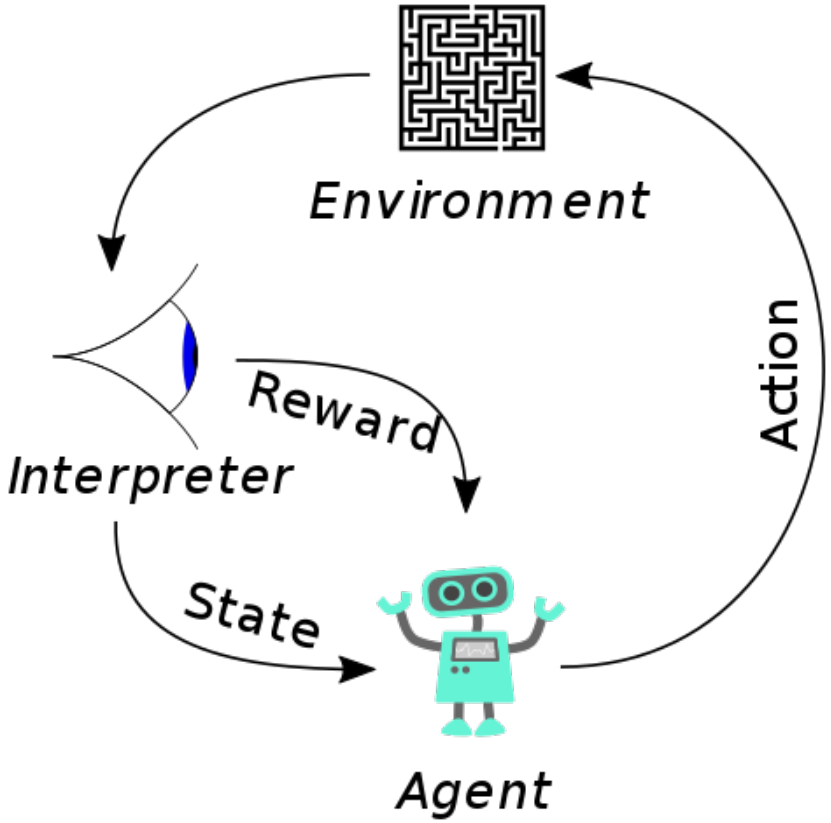
Step 4: Validation

Zoph & Le 2017

Architecture Search



Neuroevolution

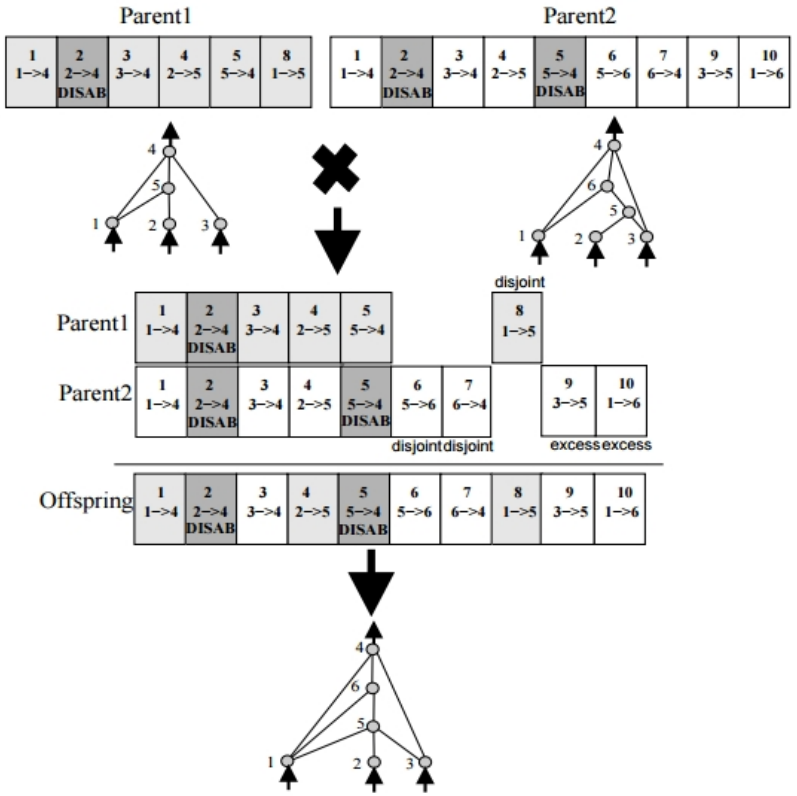


Reinforcement Learning

Wikipedia

Architecture Search

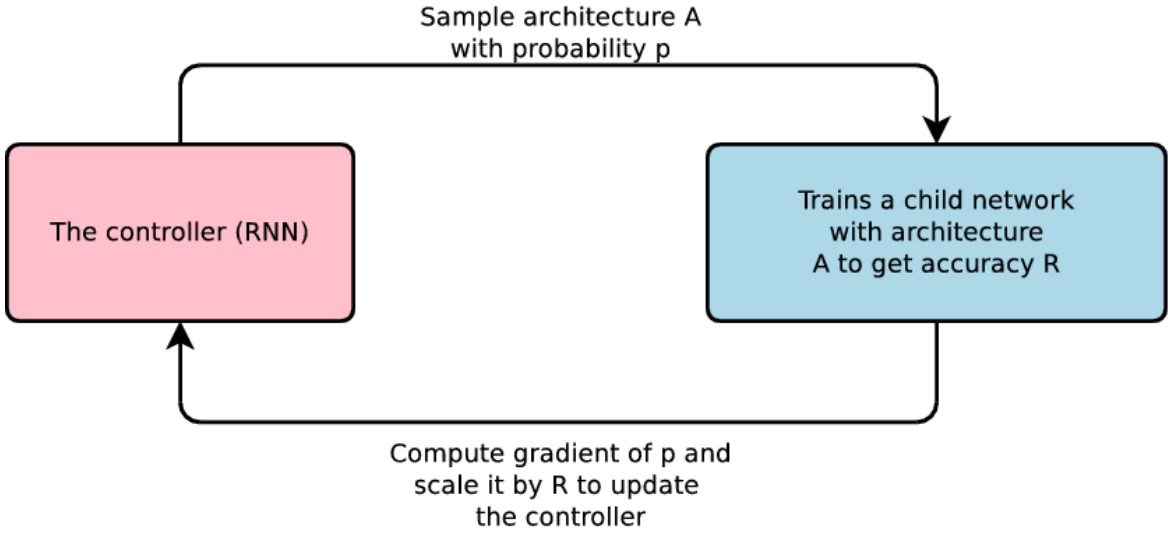
Neuroevolution



Crossover
Mutate
Speciation
Selection by Fitness or Novelty

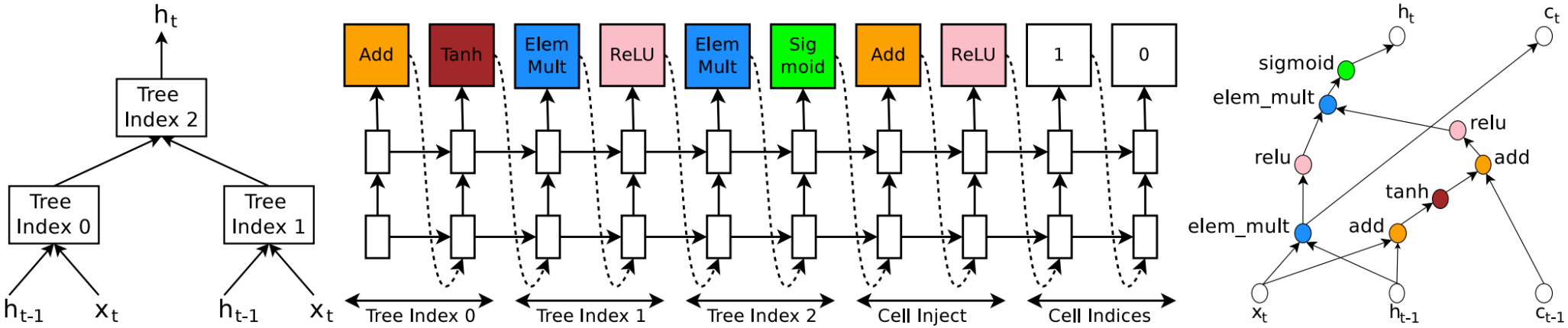
NEAT, Stanley 2002

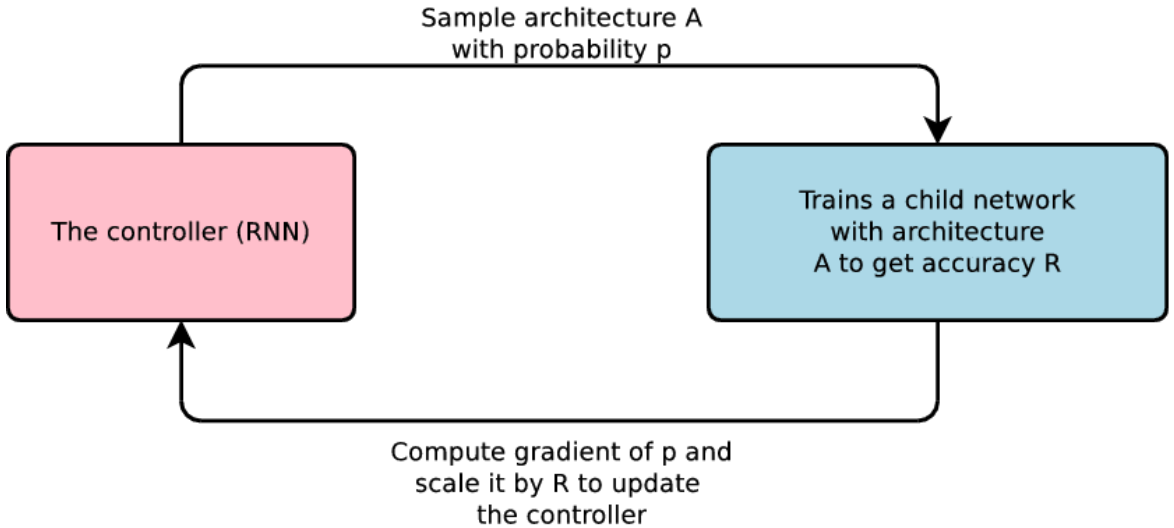
Reinforcement Learning



Networks that generate Networks

Controller networks generate architecture

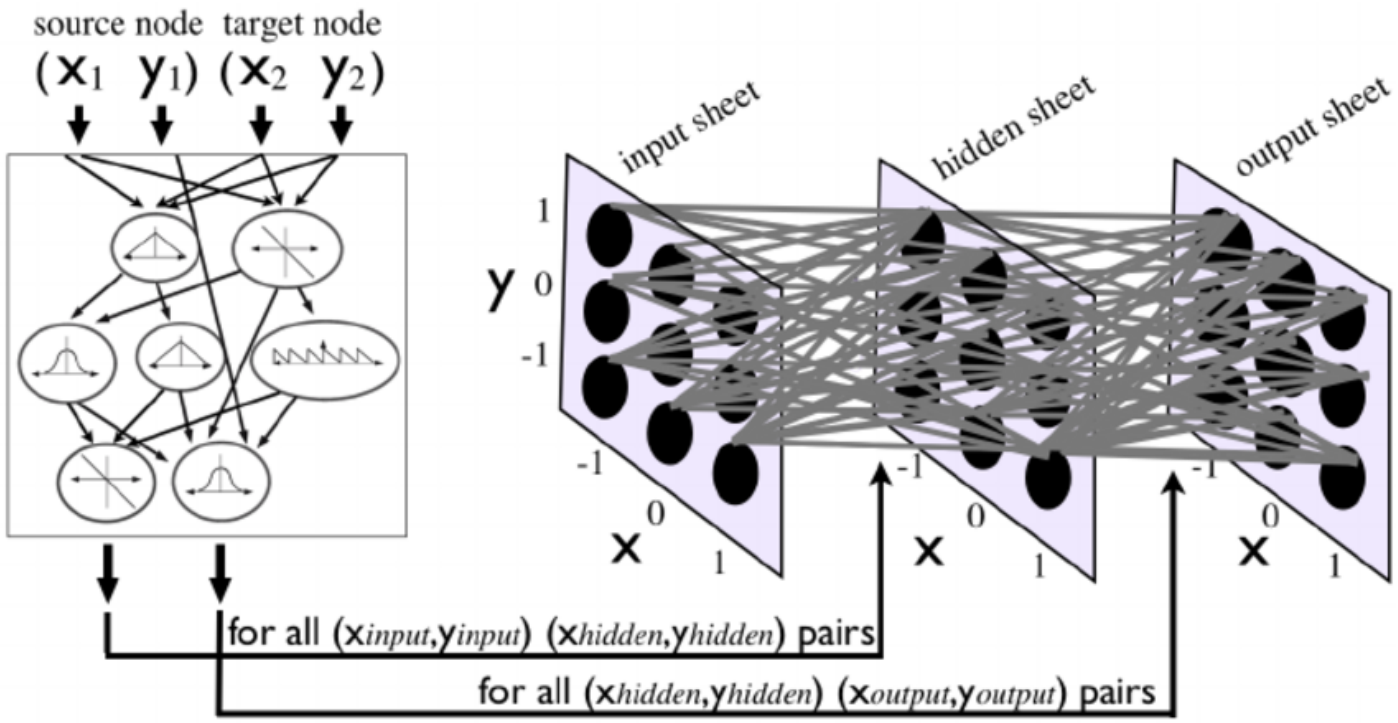




Zoph & Le 2017

Networks that generate Networks

Hypernetworks generate weights

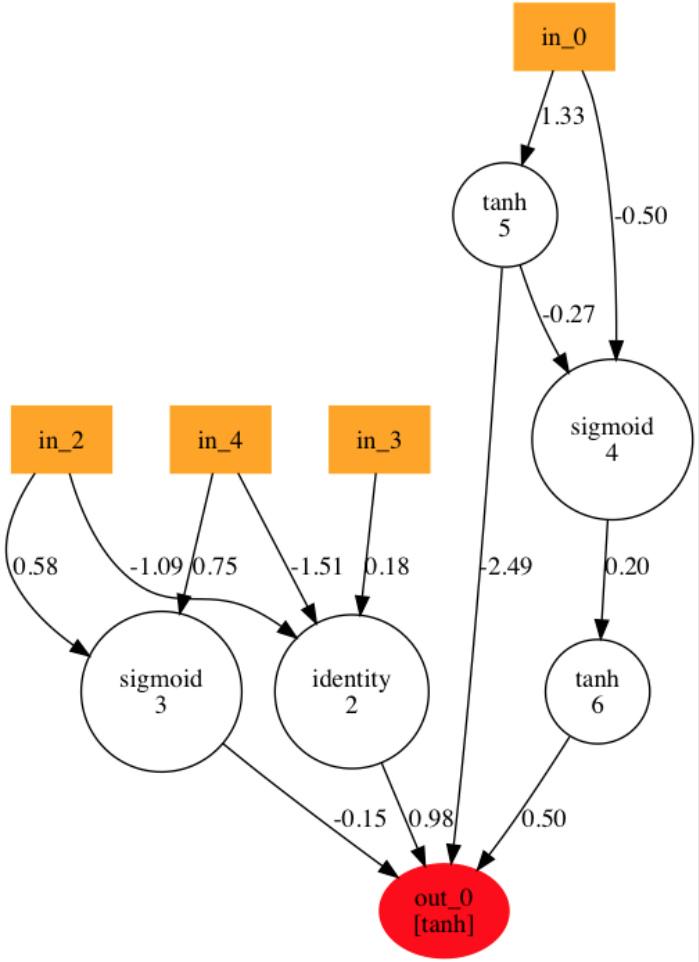


HyperNEAT

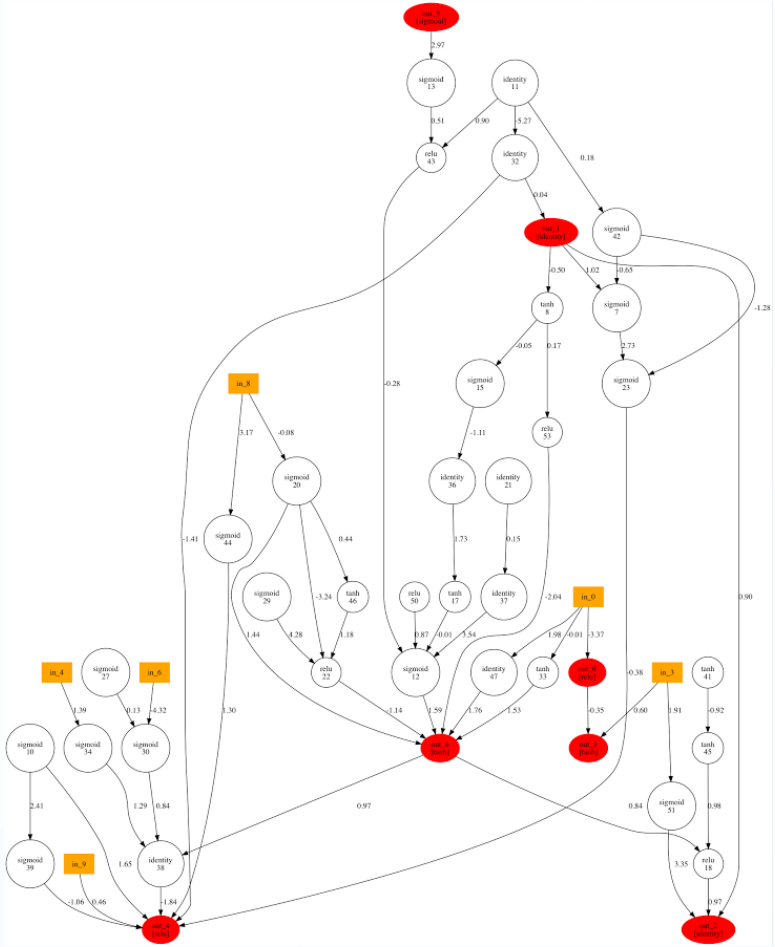
Ken Stanley 2007

Darwin Architecture Outputs

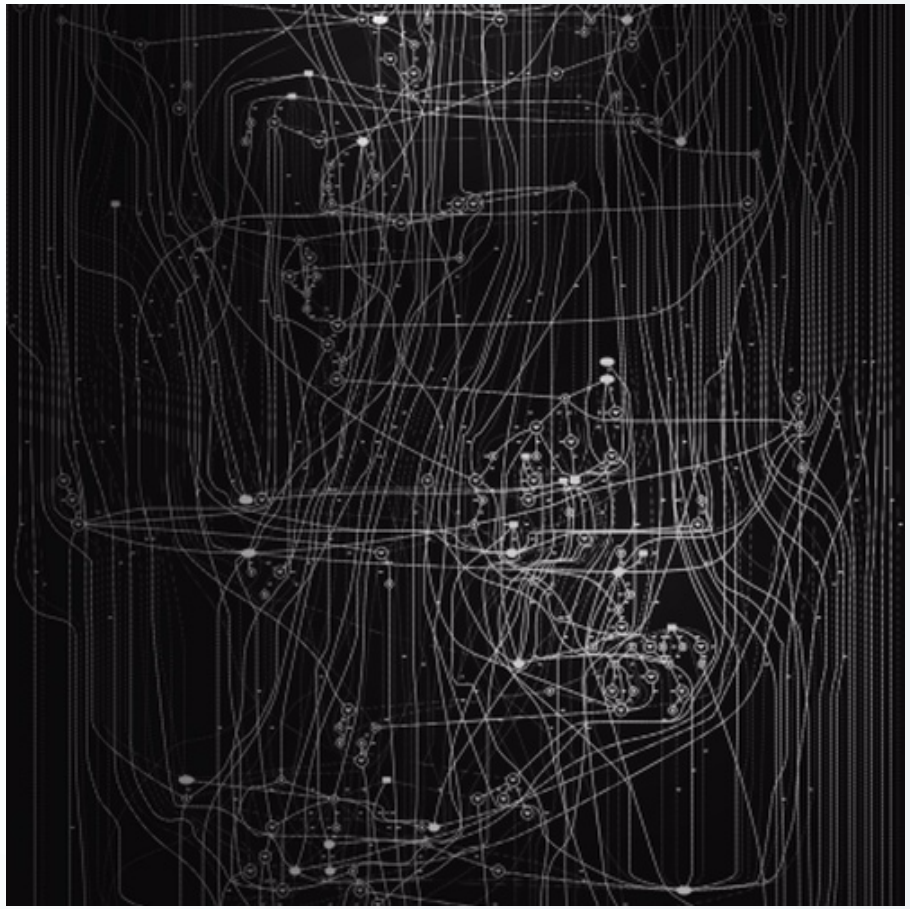
Predicts bearing failure
(400 generations)

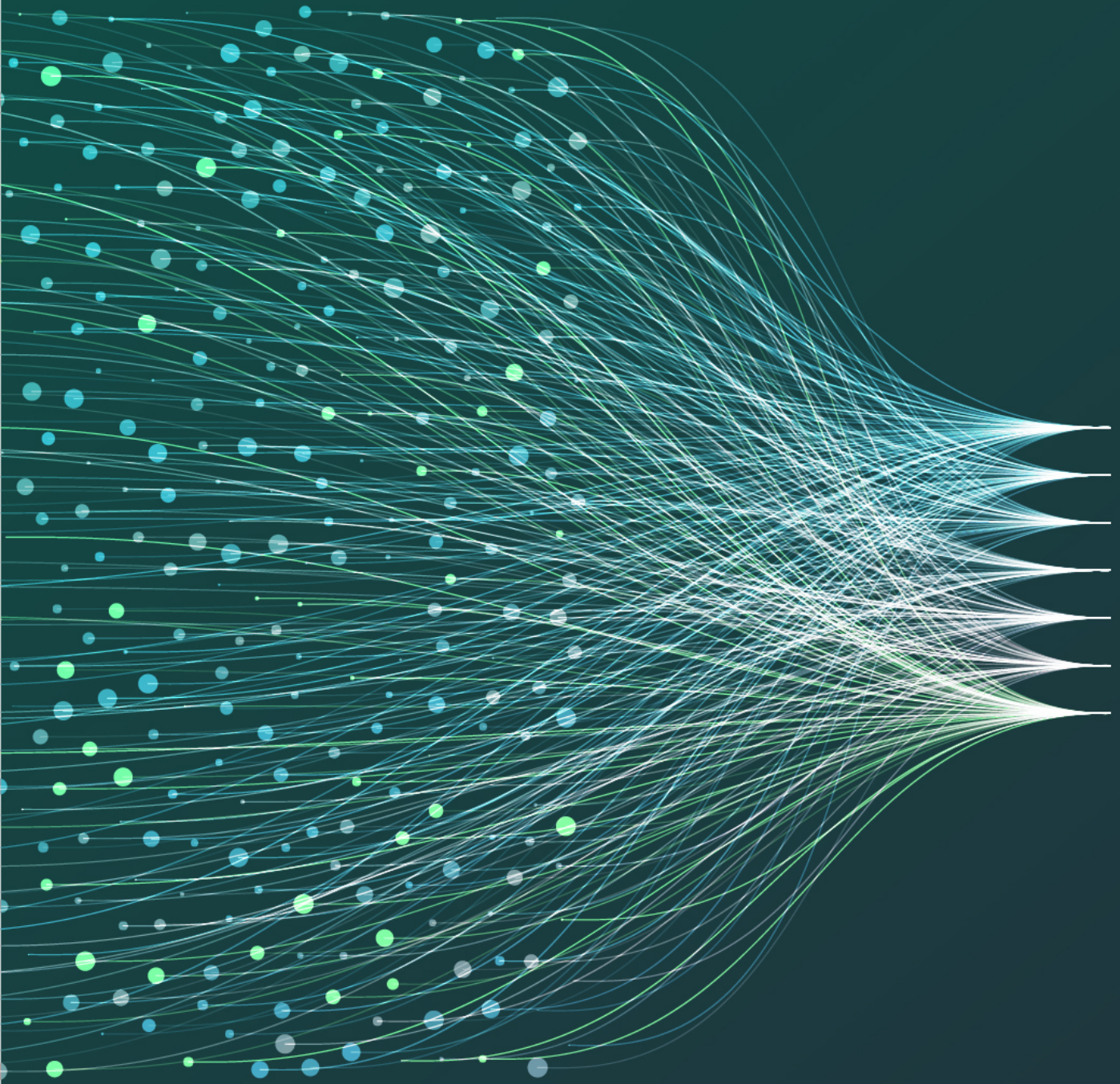


Flies an Airplane
(900 generations)



Trades Dow Futures
(1000s of generations)



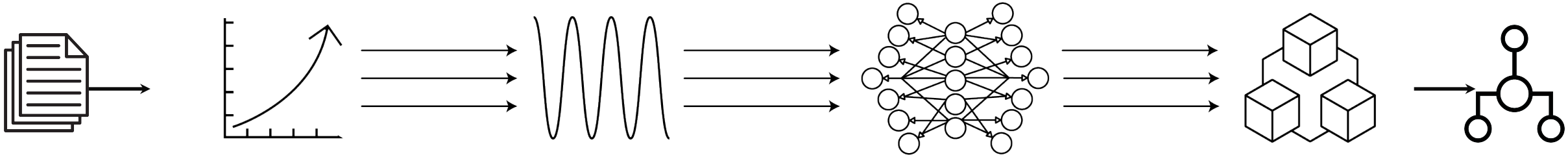


sparkcognition™

Darwin

Darwin is a product designed to automate model building in the data science process

Darwin



Structured Data

Preprocessing

- Scaling
- Imputation
- One-Hot Encoding
- Balancing

Feature Generation

- Time Domain
- Fourier Domain
- Risk Index

Feature Selection

- Raw Features
- Generated Features

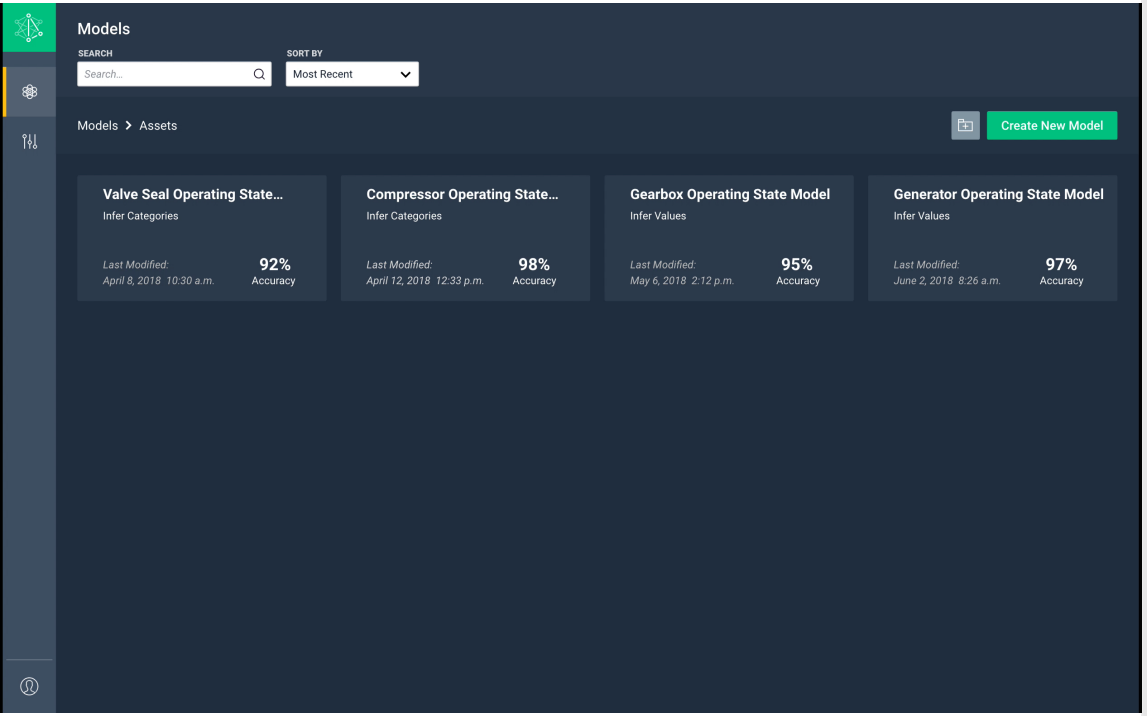
Model Building

- Classification
- Regression
- Clustering
- Anomaly Detection

Model

UI / REST API / Python

Darwin User Experience



jupyter Darwin NBM Example Last Checkpoint: 10/15/2018 (autosaved)

Logout

File Edit View Insert Cell Kernel Widgets Help

Trusted | Python 3

Run | Markdown

Predictions

Perform model prediction on the the training dataset.

In [69]:
```
status, artifact = ds.run_model(dataset_name, model)
sleep(1)
ds.wait_for_job(artifact['job_name'])
```

```
{'status': 'Running', 'starttime': '2018-10-15T11:40:50.185599', 'endtime': None, 'percent_complete': 0, 'job_type':
'RunModel', 'loss': 0.10883557872894482, 'generations': 6, 'dataset_names': ['wind_turbine.csv'], 'artifact_names':
['ee746c8870da47a0b85604eb2c5add31'], 'model_name': 'nbm_model', 'job_error': ''}
{'status': 'Running', 'starttime': '2018-10-15T11:40:50.185599', 'endtime': None, 'percent_complete': 0, 'job_type':
'RunModel', 'loss': 0.10883557872894482, 'generations': 6, 'dataset_names': ['wind_turbine.csv'], 'artifact_names':
['ee746c8870da47a0b85604eb2c5add31'], 'model_name': 'nbm_model', 'job_error': ''}
{'status': 'Complete', 'starttime': '2018-10-15T11:40:50.185599', 'endtime': '2018-10-15T11:41:10.459713', 'percent_c
omplete': 100, 'job_type': 'RunModel', 'loss': 0.10883557872894482, 'generations': 6, 'dataset_names': ['wind_turbin
e.csv'], 'artifact_names': ['ee746c8870da47a0b85604eb2c5add31'], 'model_name': 'nbm_model', 'job_error': ''}
```

Out[69]:
```
(True, 'Job completed')
```

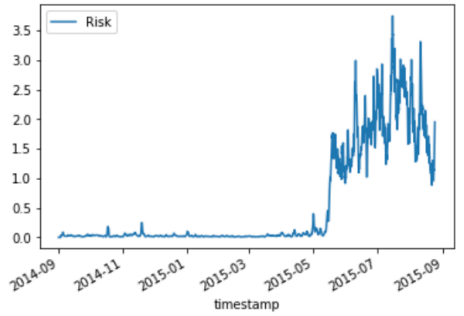
Download predictions from Darwin's server.

In [70]:
```
status, prediction = ds.download_artifact(artifact['artifact_name'])
```

Create plots showing the risk index prediction of the model

In [71]:
```
#Plot the risk predictions
prediction.set_index(pd.to_datetime(df['timestamp']), inplace=True)
prediction.plot()
```

Out[71]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a28546978>
```



https://www.sparkcognition.com/darwin-trial/