Sparse Neural Computation

Dr Varun Ojha
Department of Computer Science
University of Reading

at

Centre for Computational Science and Mathematical Modelling
Coventry University
10 June 2022

Intrinsic Intelligence of a child’s mind

Slide inspiration: Josh Tenenbaum, Prof. MIT, USA
Video Source:
https://www.youtube.com/watch?v=dEnDjyWHN4A
(Accessed on 21 Feb 2021)

Learning

Training the *Mind* of Species

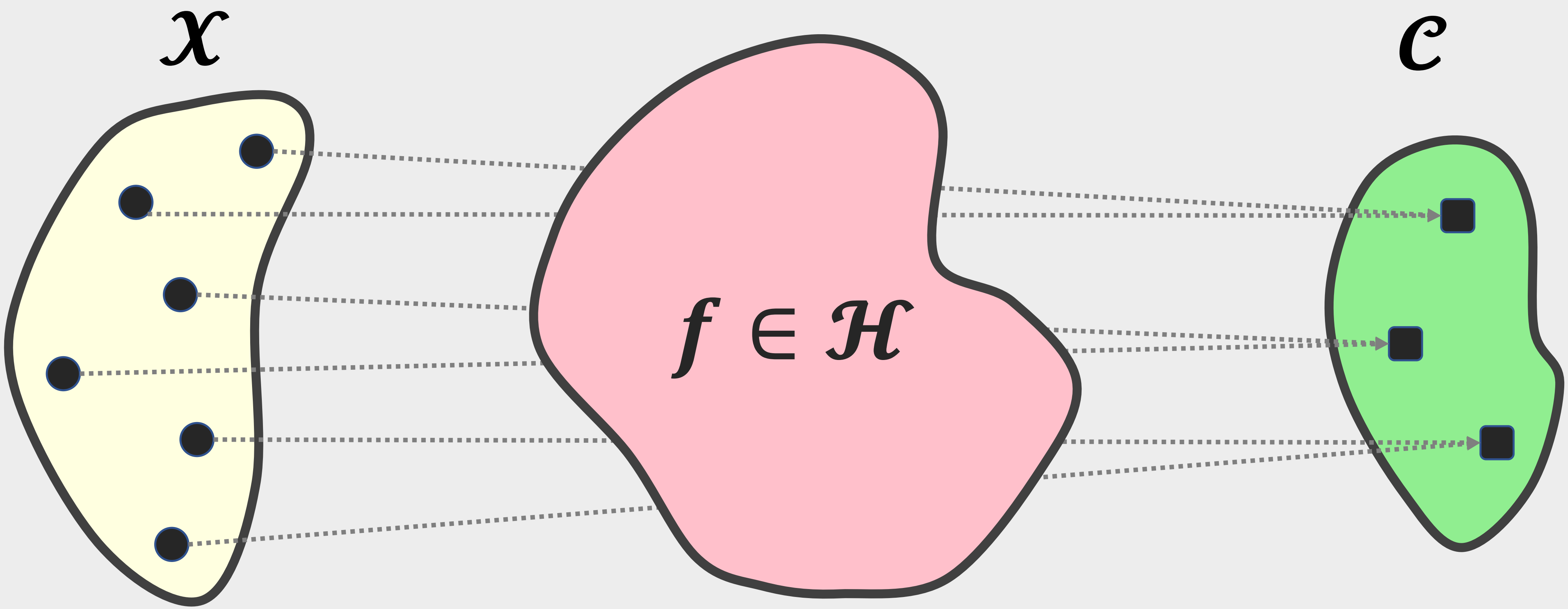
Video source:

https://www.youtube.com/watch?v=nbrTOcUnjNY

(Accessed on 21 Feb 2021)

Learning $f: X \rightarrow y$

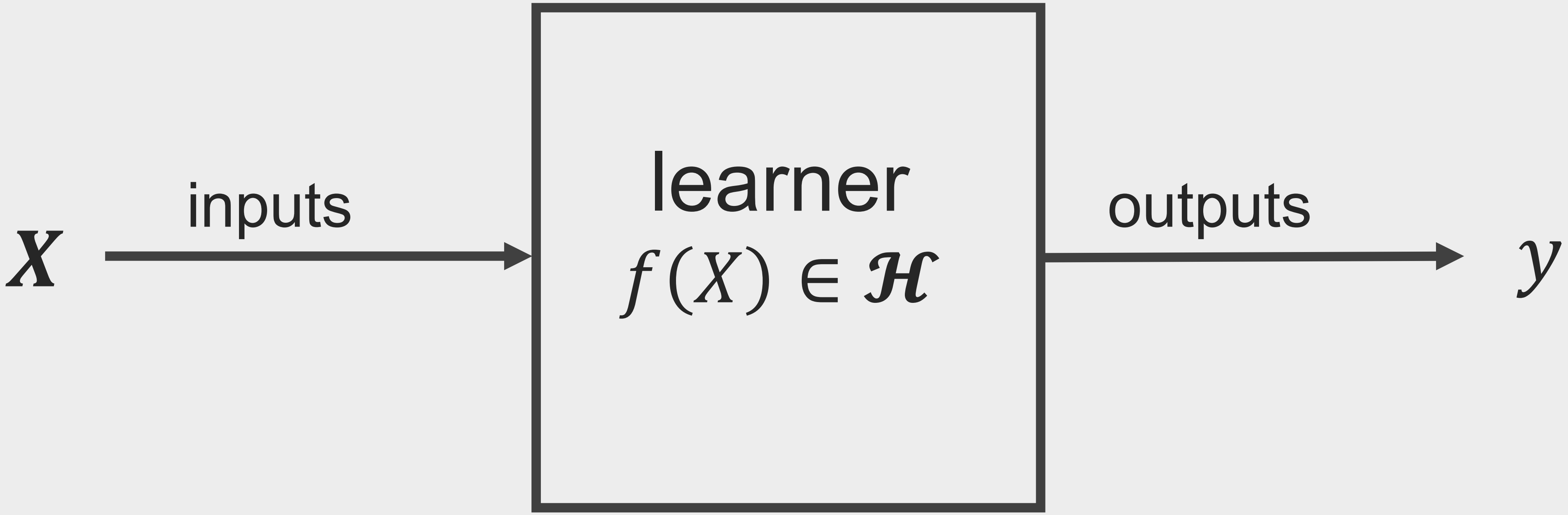
Find the <u>unknown</u> target function f that does the mapping



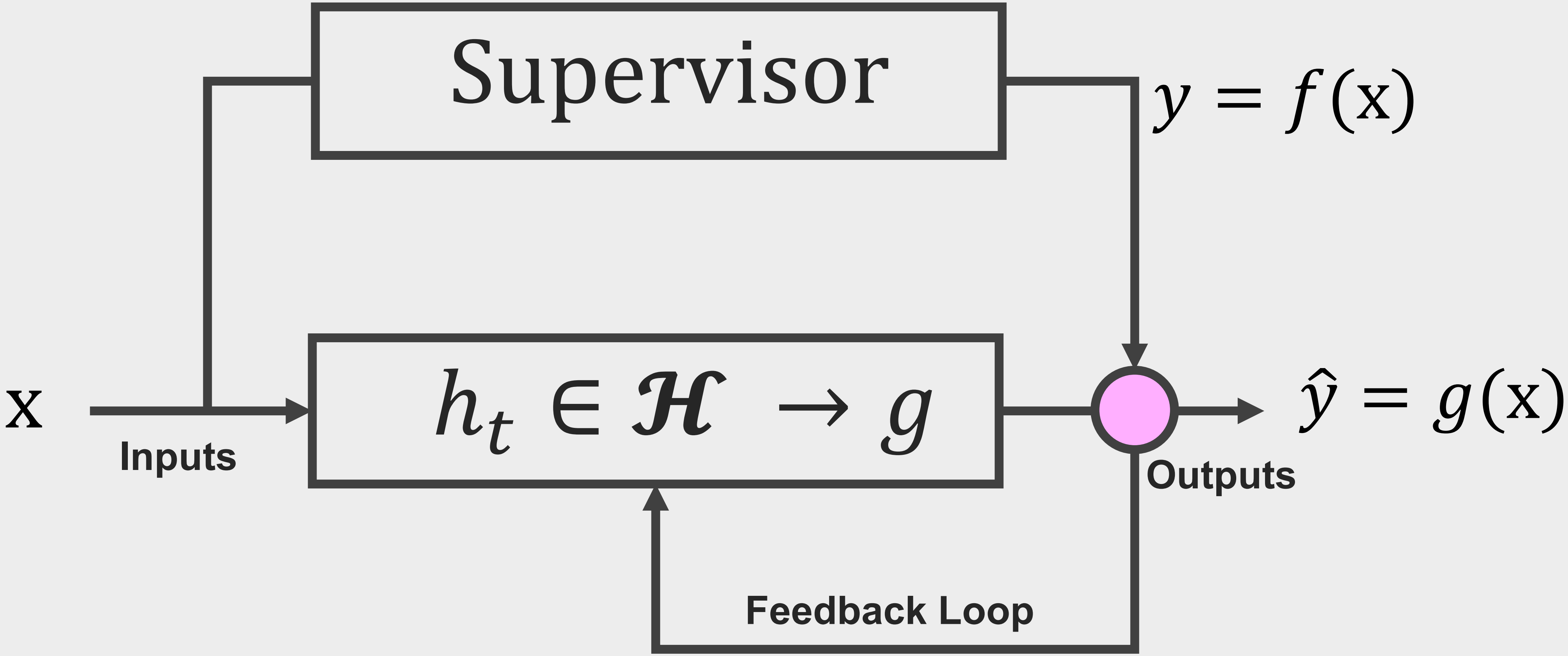
Inputs $\mathbf{X} \in$ Input space $\mathcal{X}$ hypothesis space $\mathcal{H}$ outputs $\boldsymbol{y} \in$ concept space $\mathcal{C}$

Learning $\boldsymbol{f}: \boldsymbol{X} \rightarrow \boldsymbol{y}$

Supervised learning approximates a function $\mathbf{g} \sim \boldsymbol{f}$ for mapping inputs $\boldsymbol{X}$ to outputs $\boldsymbol{y}$



How to Produces the Function $g: X \rightarrow y$



What Learning Needs

Learning needs the method(s) to

Represent

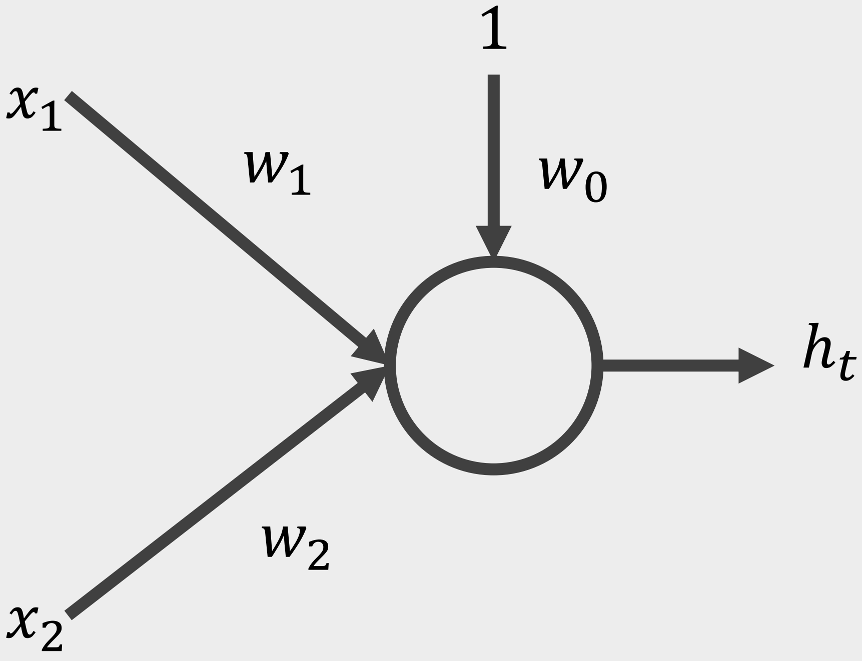
Evaluate

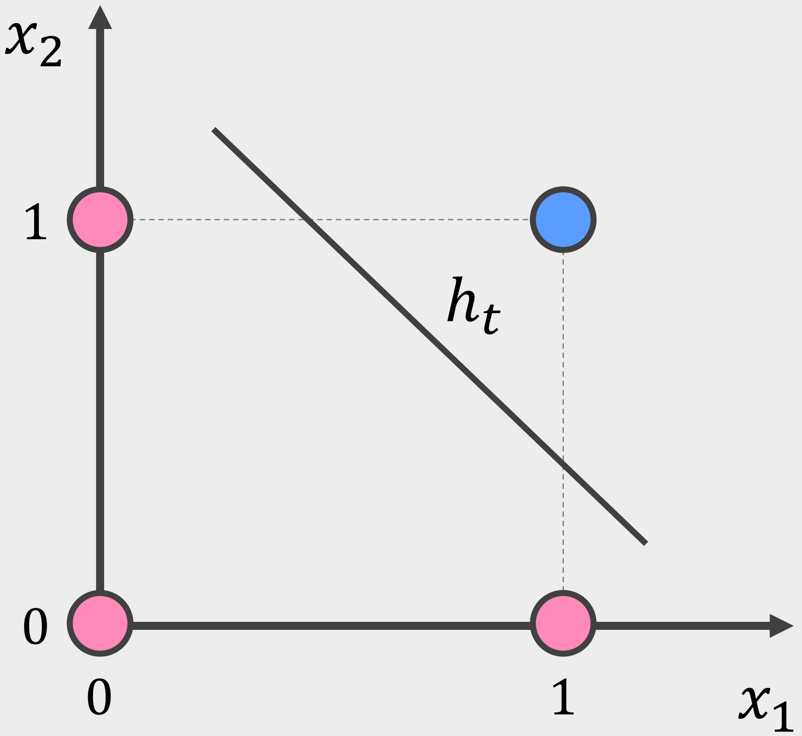
Optimize

a hypothesis h_t

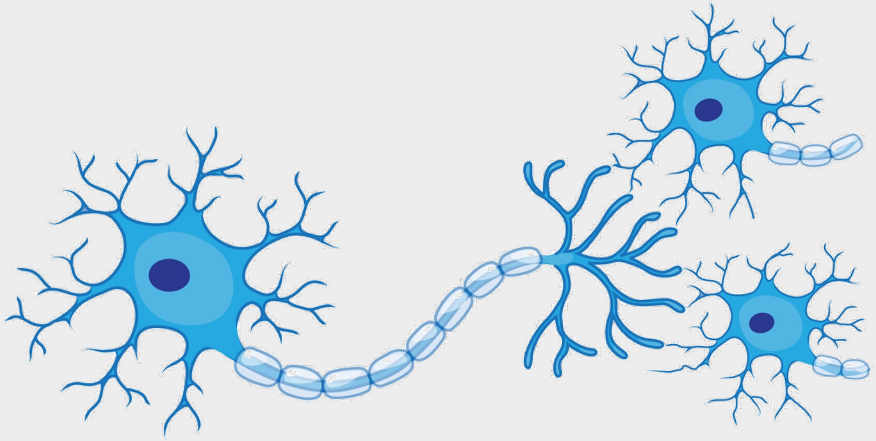
How to Represent a Hypothesis $h_t \in H$

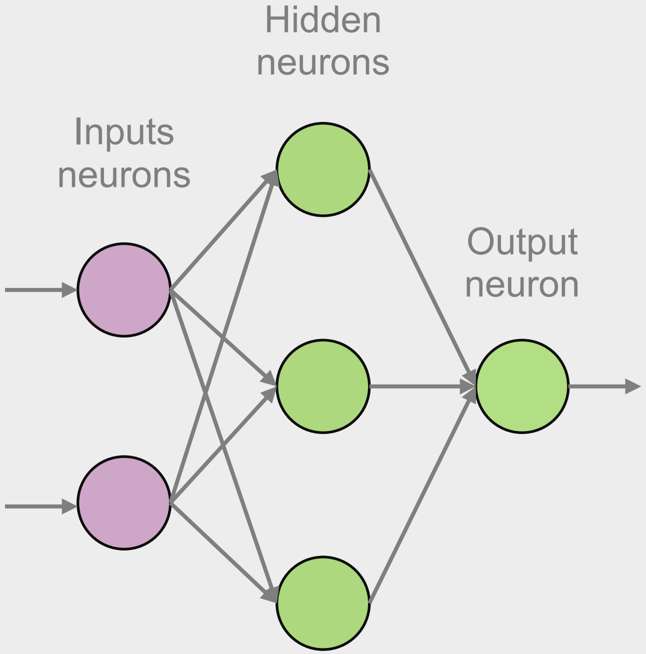
A line separating data can be considered a hypothesis

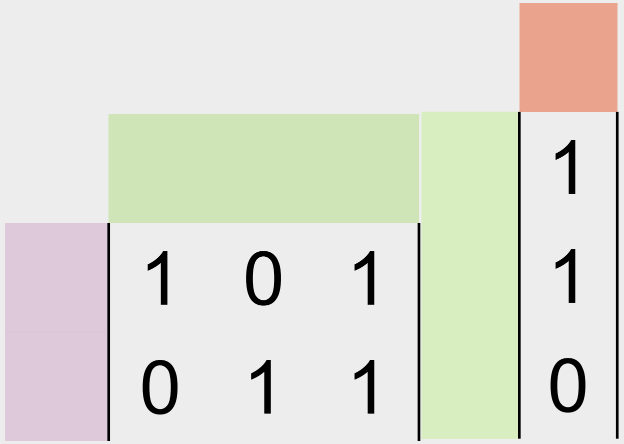




Learning Systems: Neural Networks







1 Biological networks of neurons in human brains

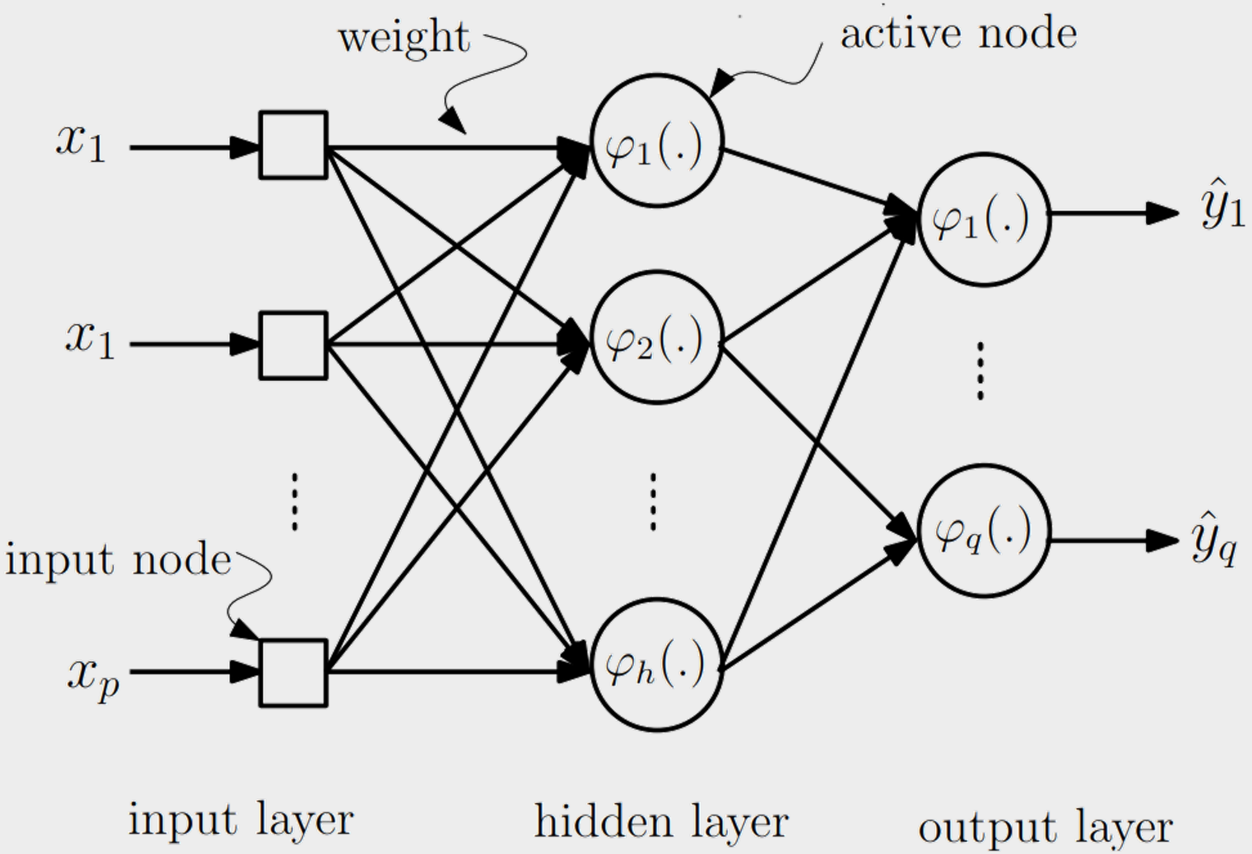
2 AI representation of biological neural networks

3 Mathematical representation of the neural networks

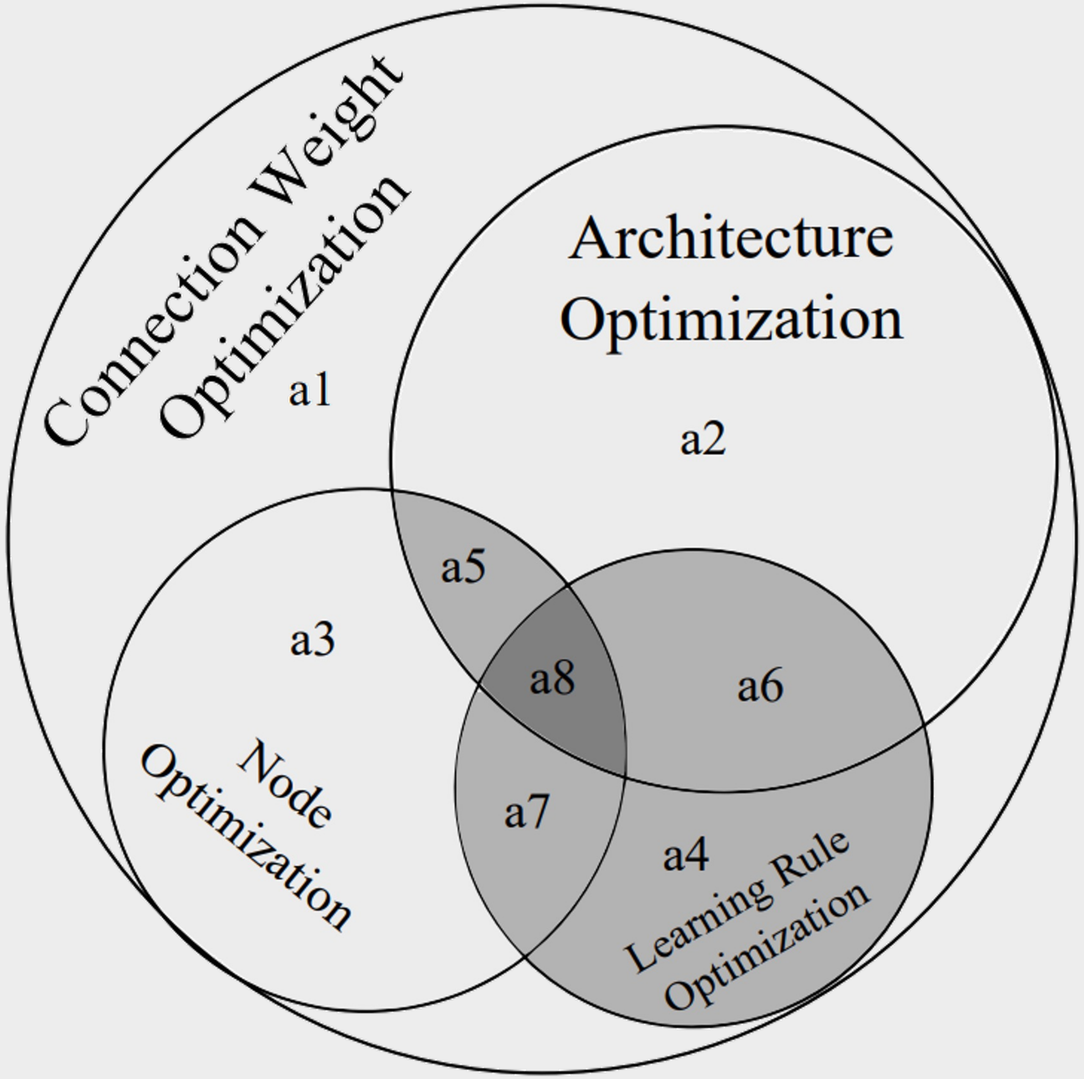
Neural Networks

NN components:

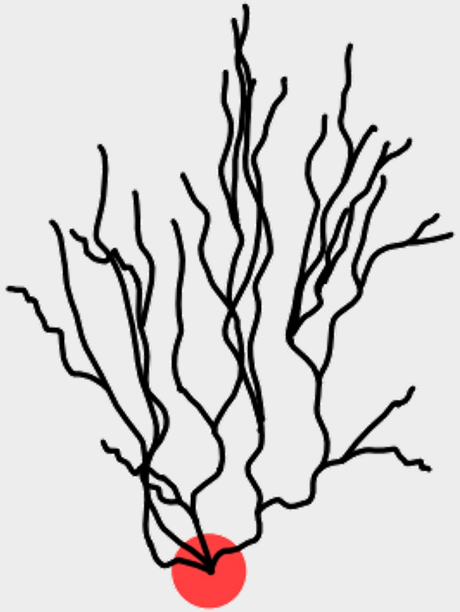
- Inputs
- Weights
- Architecture
- Activation functions
- Learning algorithms



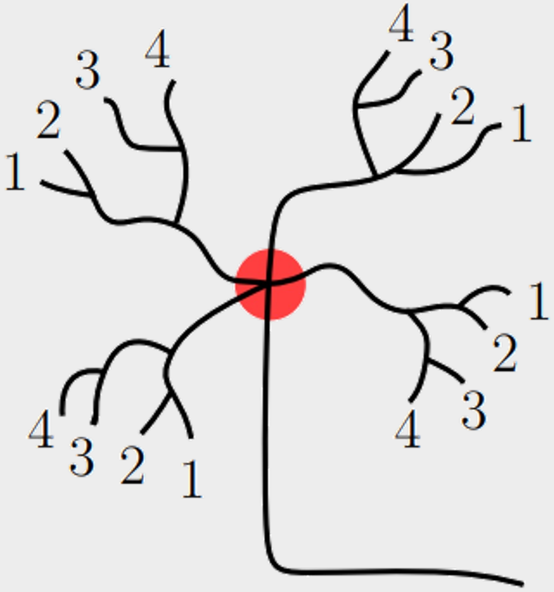
What Could be optimized?



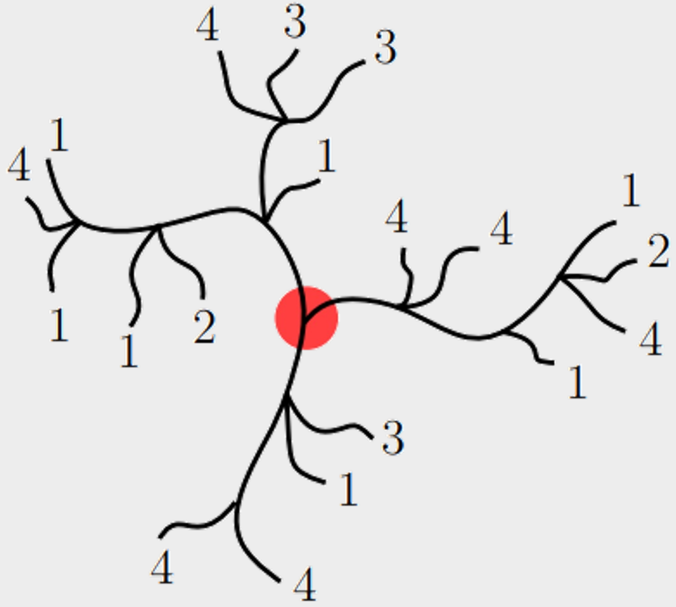
Plausible Biological Inspiration



Travis et al. (2005)



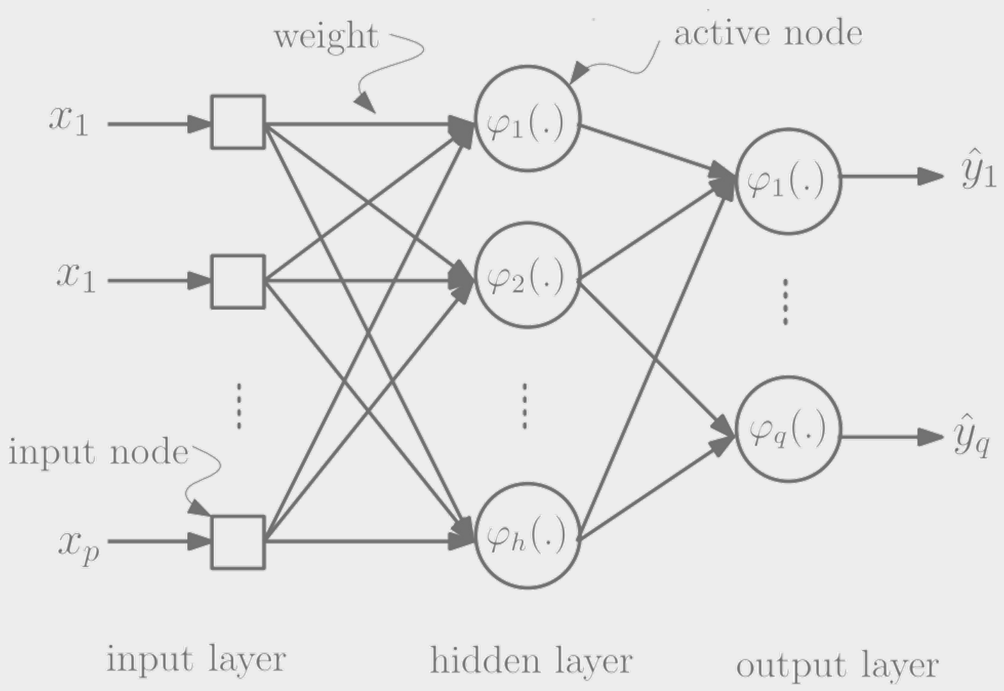
Jones and Kording (2021)

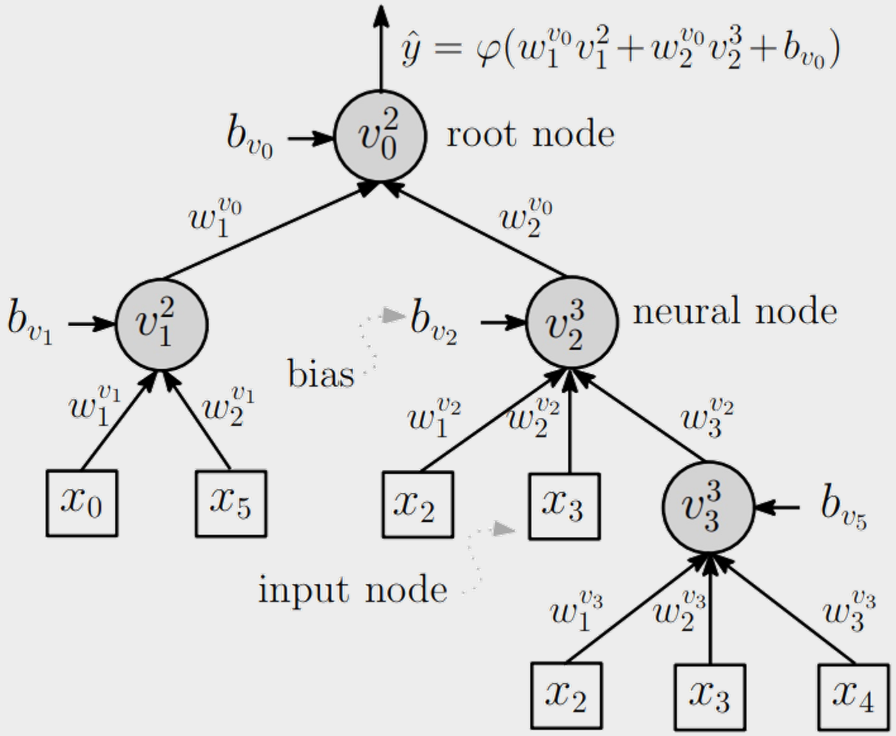


Ojha and Nicosia (2022)

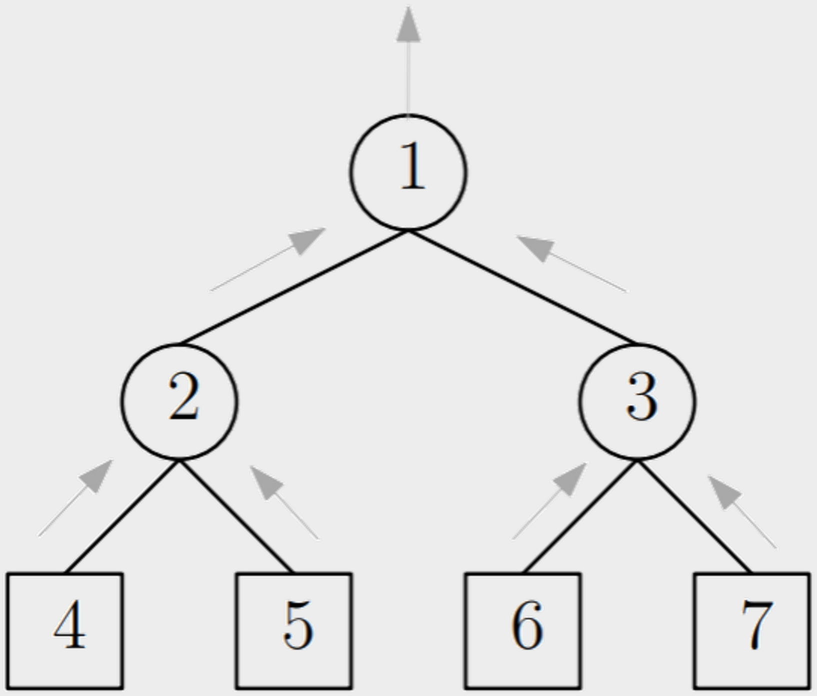
Neural Tree

Neural Networks Architecture Search





Neural Computation



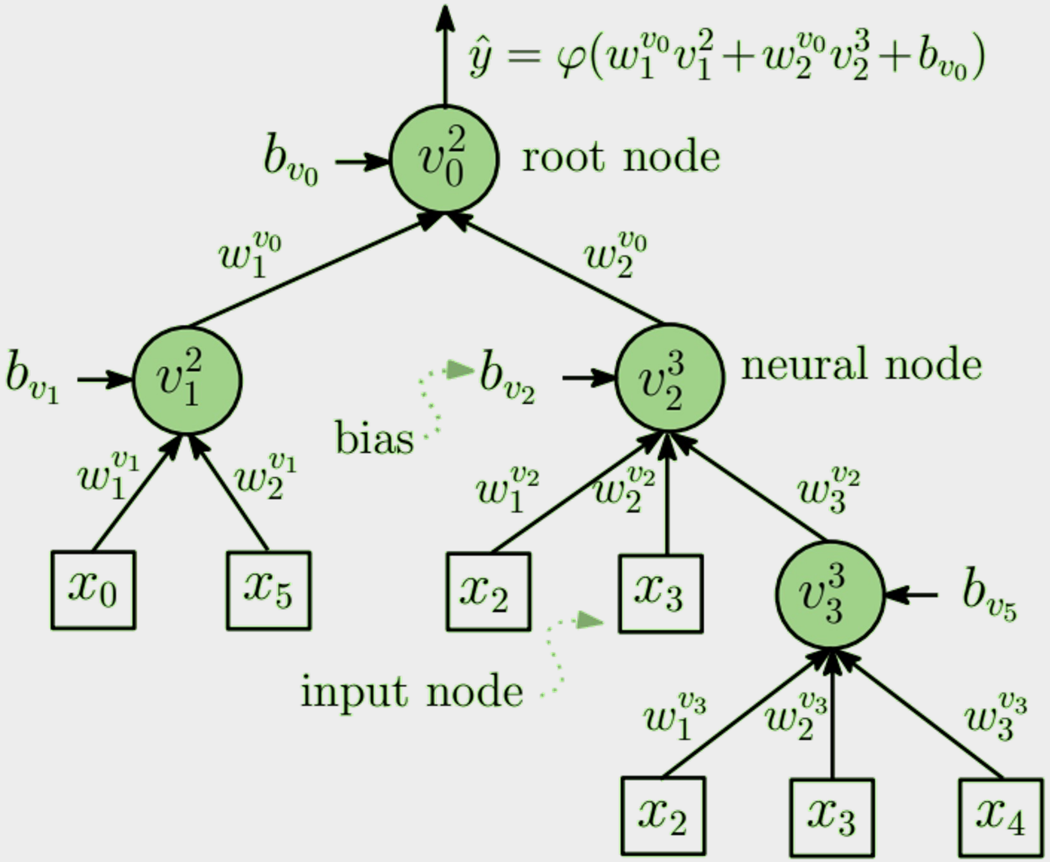
$$[((4\ 5) \rightarrow 2) \quad ((6\ 7) \rightarrow 3)] \rightarrow 1$$

forward pass: post-order

Ojha et al (2022), Neural Networks

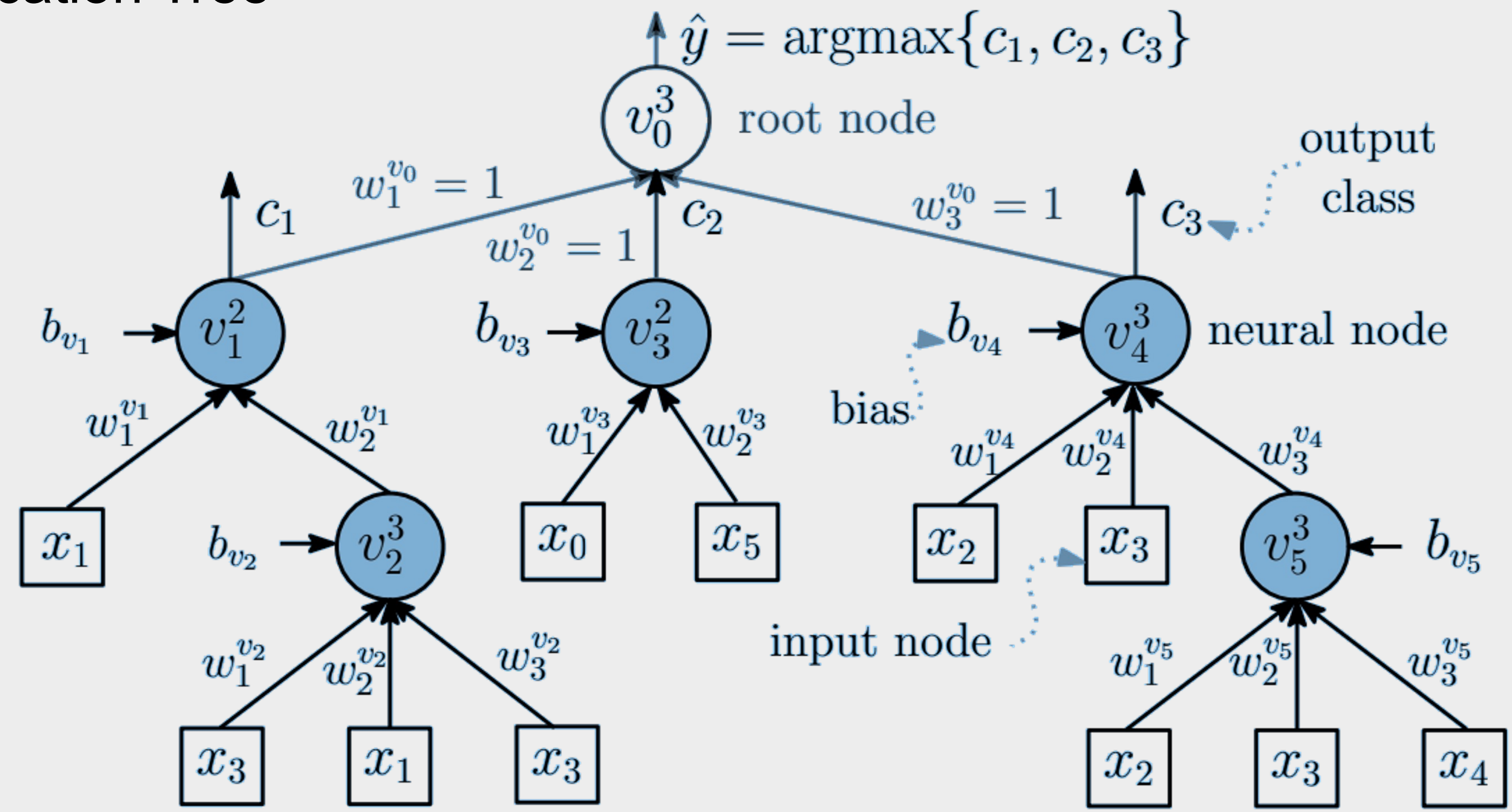
Types of Neural Tree

Regression Tree



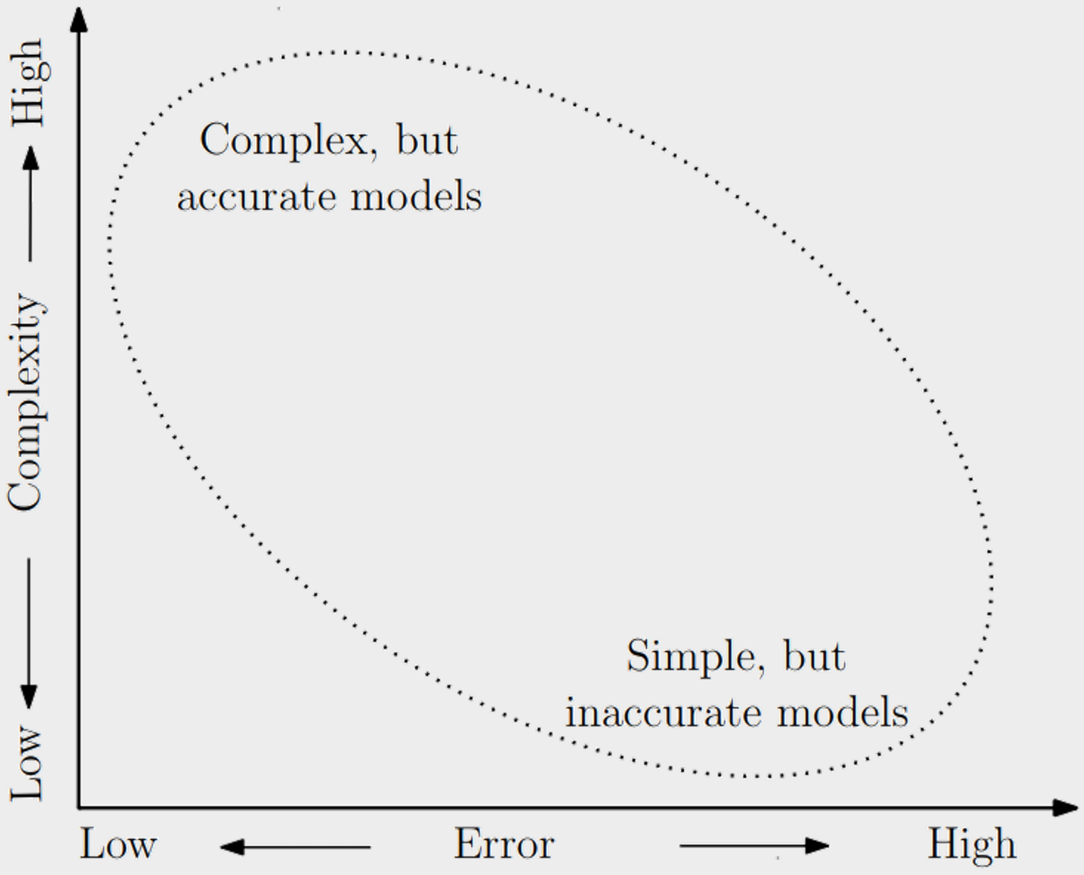
Types of Neural Tree

Classification Tree



Neural Architecture Search

Trade-offs

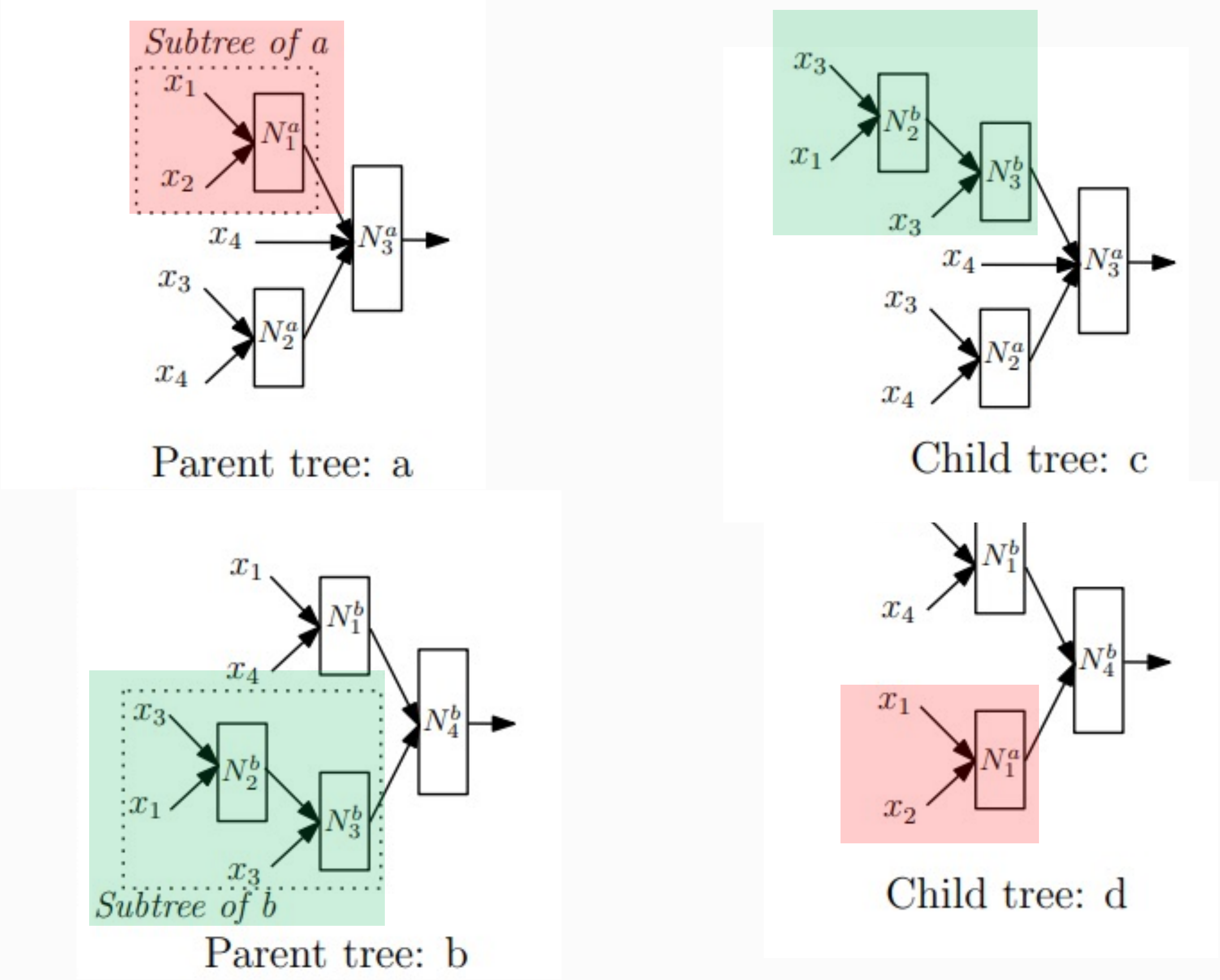


Neural Architecture Search

Trade-offs

Multiobjective Genetic Programming Crossover

Ojha et al (2017), *IEEE Trans. Fuzzy Systems*

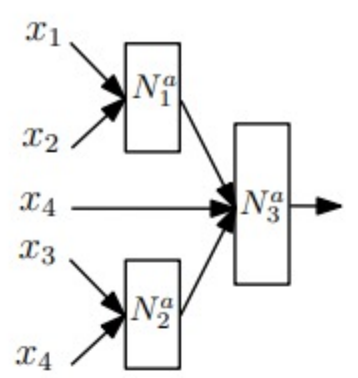


Neural Architecture Search

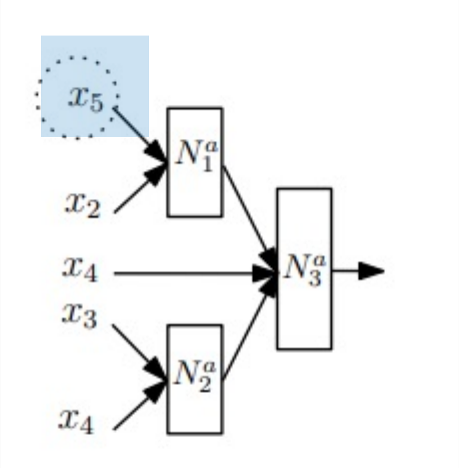
Trade-offs

Multiobjective
Genetic Programming
Mutation

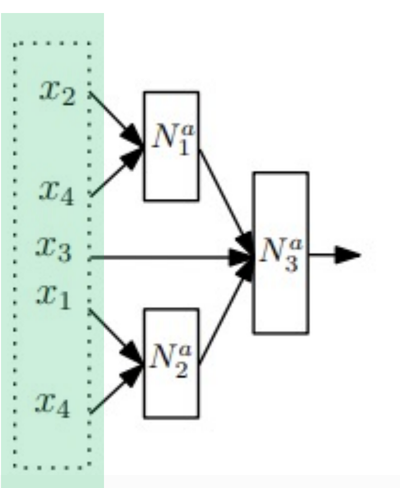
Ojha et al (2017), *IEEE Trans. Fuzzy Systems*



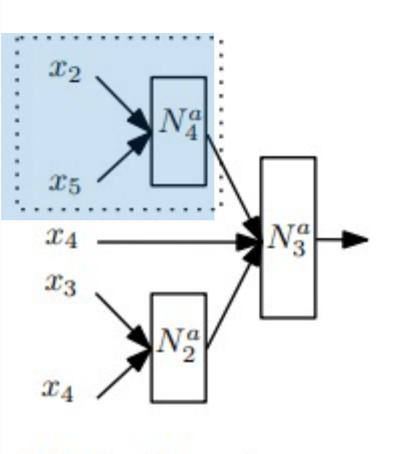
Parent tree



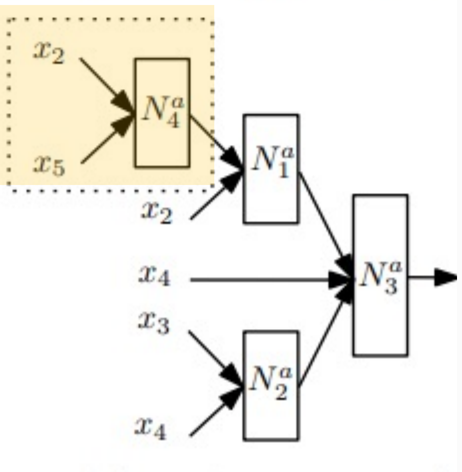
Single leaf mutation



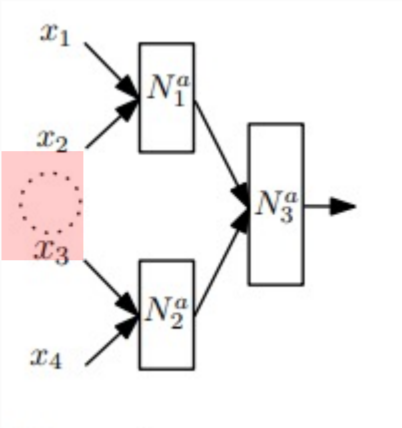
All leaves mutation



A subtree replacement



A subtree insertion

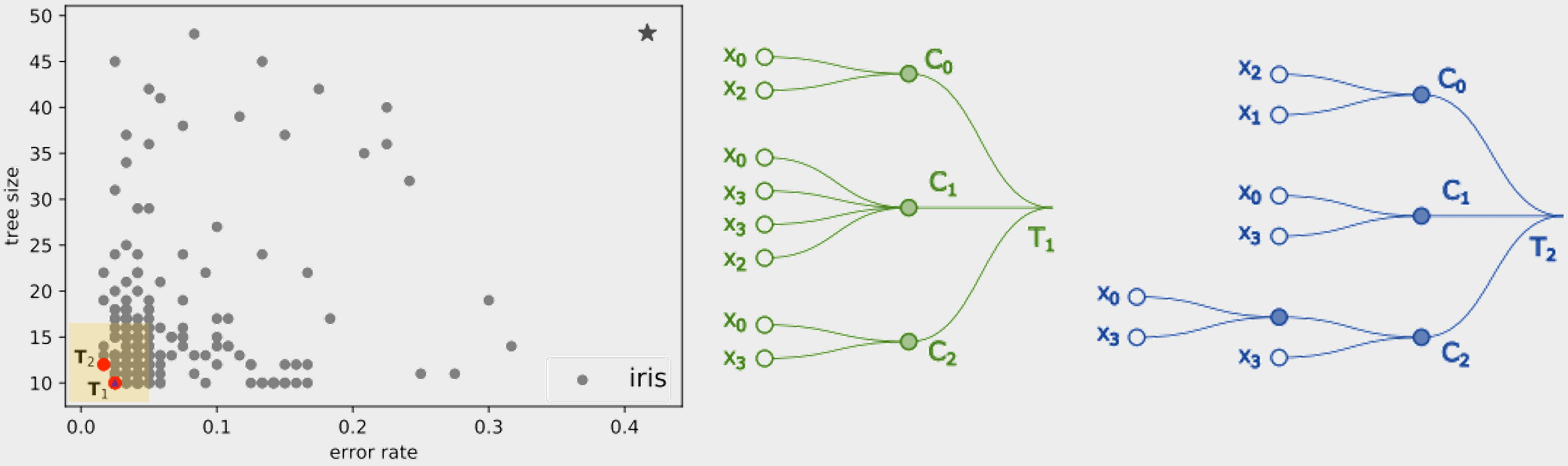


A subtree deletion

Architecture Search Trade-offs

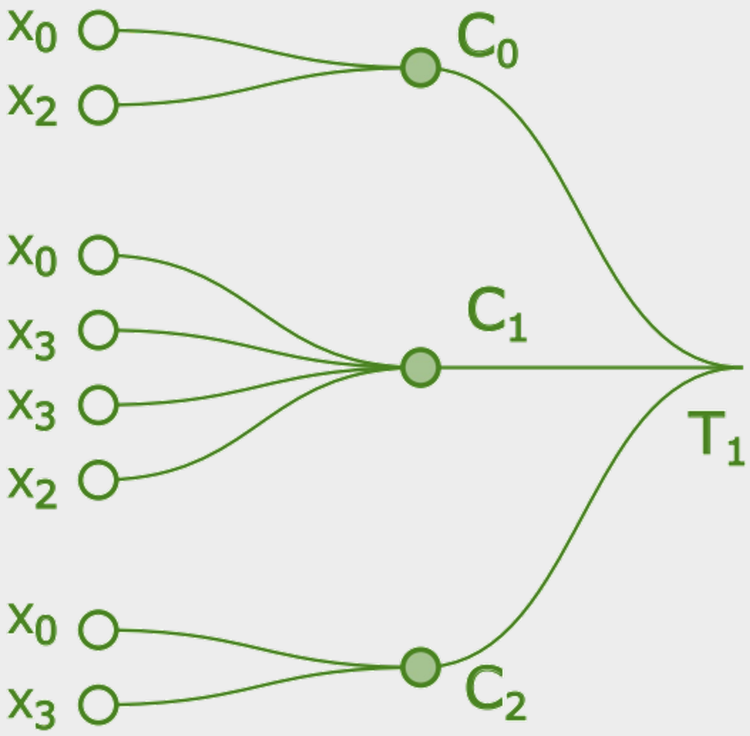
Multiobjective Genetic Programming

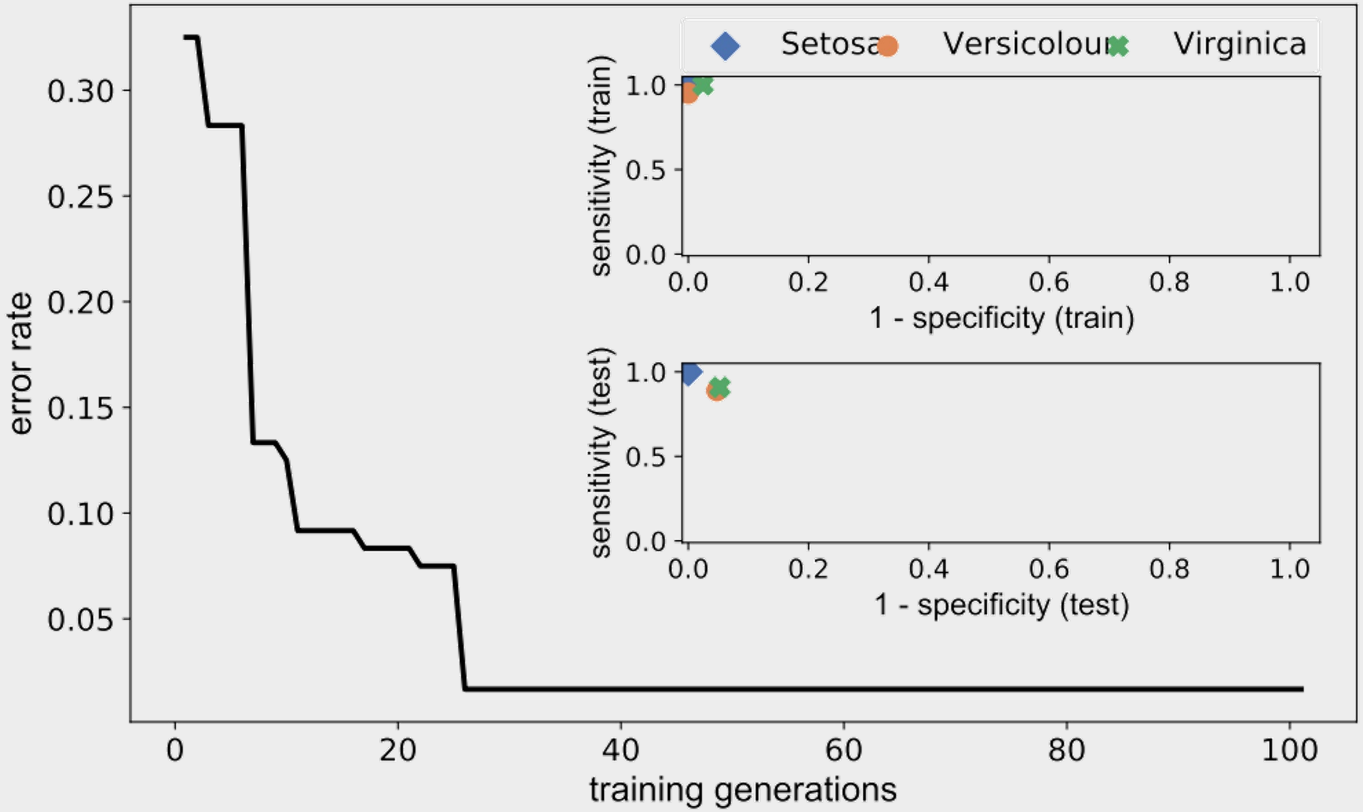
Selection of trees using Hypervolume indicator from a Pareto Front



Learnability of Classes

Competition between classes: TPR/Recall/Sensitivity vs FPR/(1 - Specificity)



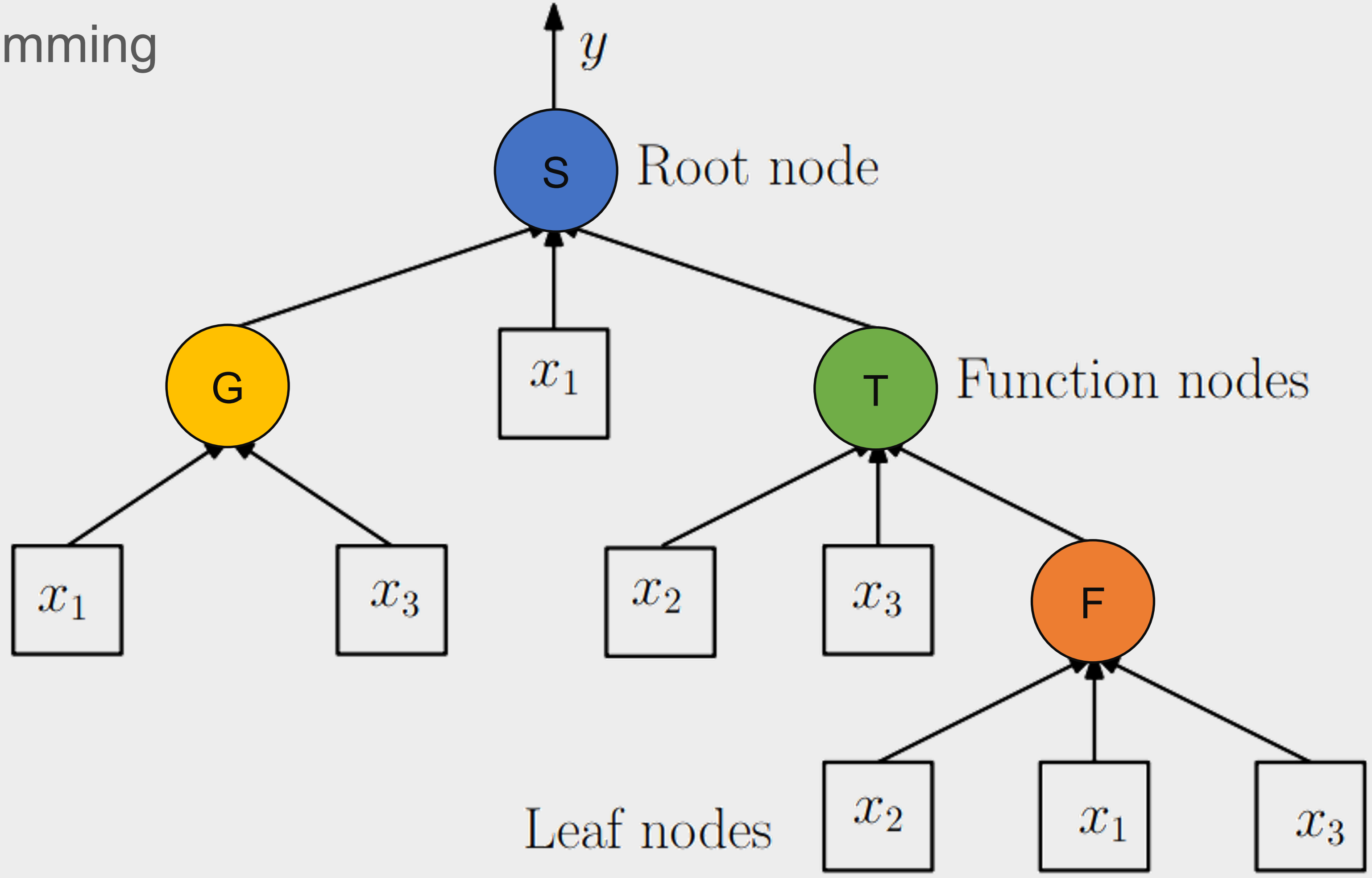


Heterogeneous Neural Tree

Multiobjective Genetic Programming

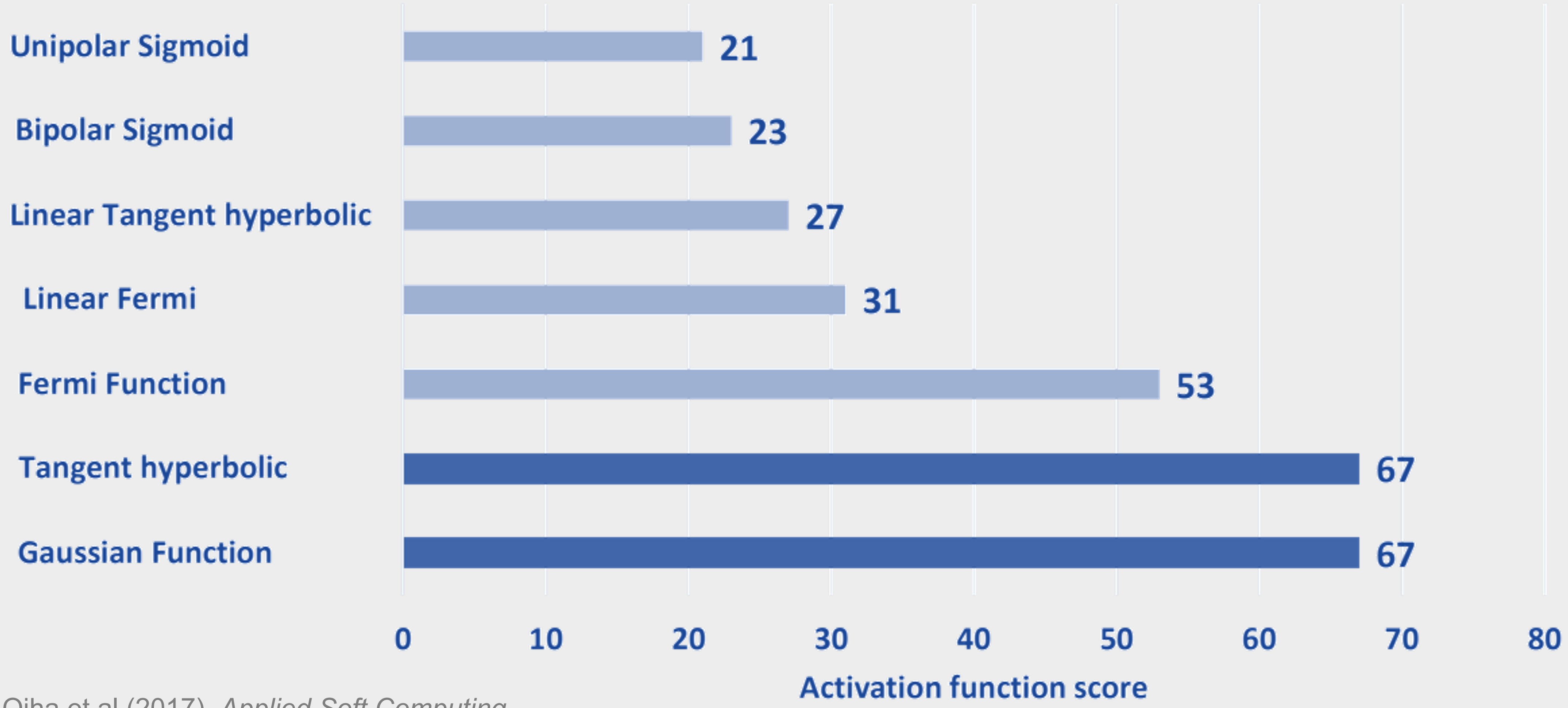
Activation Function Search

- S – Sigmoid
- G – Gaussian
- T – Tanh
- F – Fermi



Activation Function Performance

Higher values are better



Ojha et al (2017), *Applied Soft Computing*

Backpropagation Neural Tree

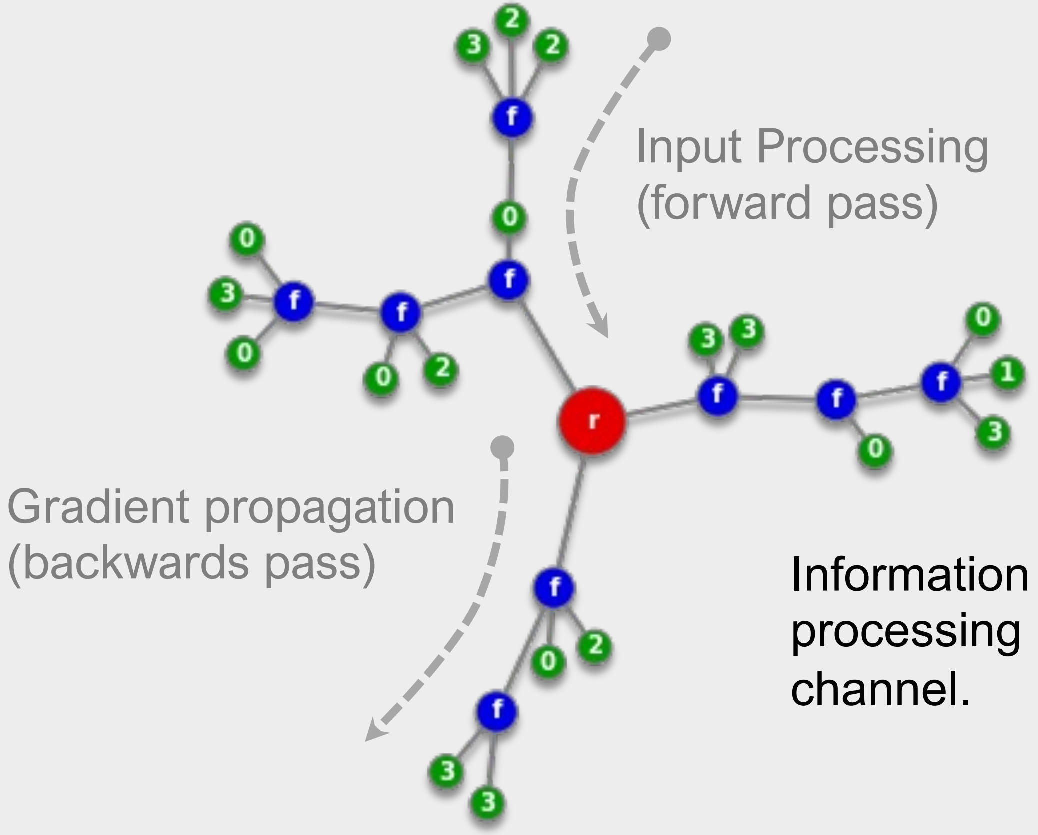
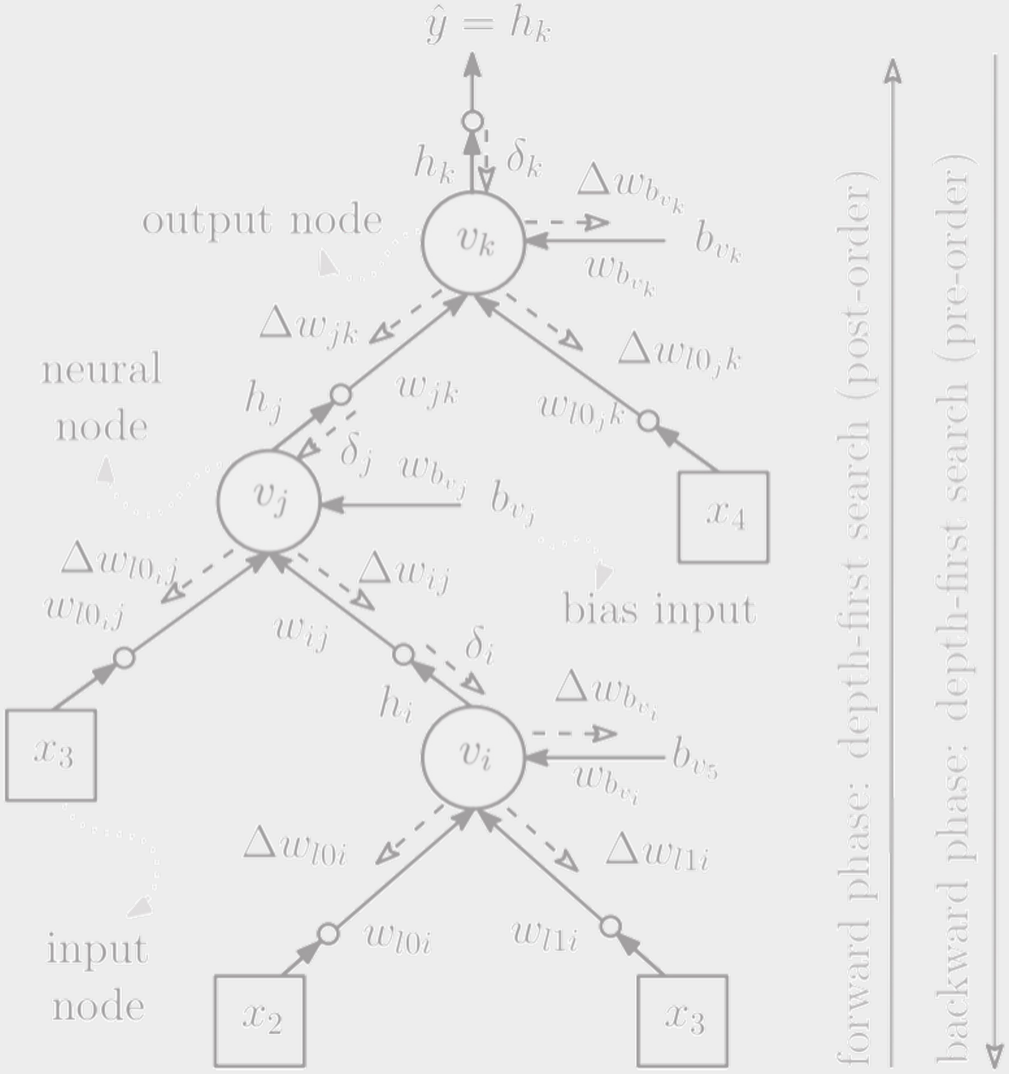
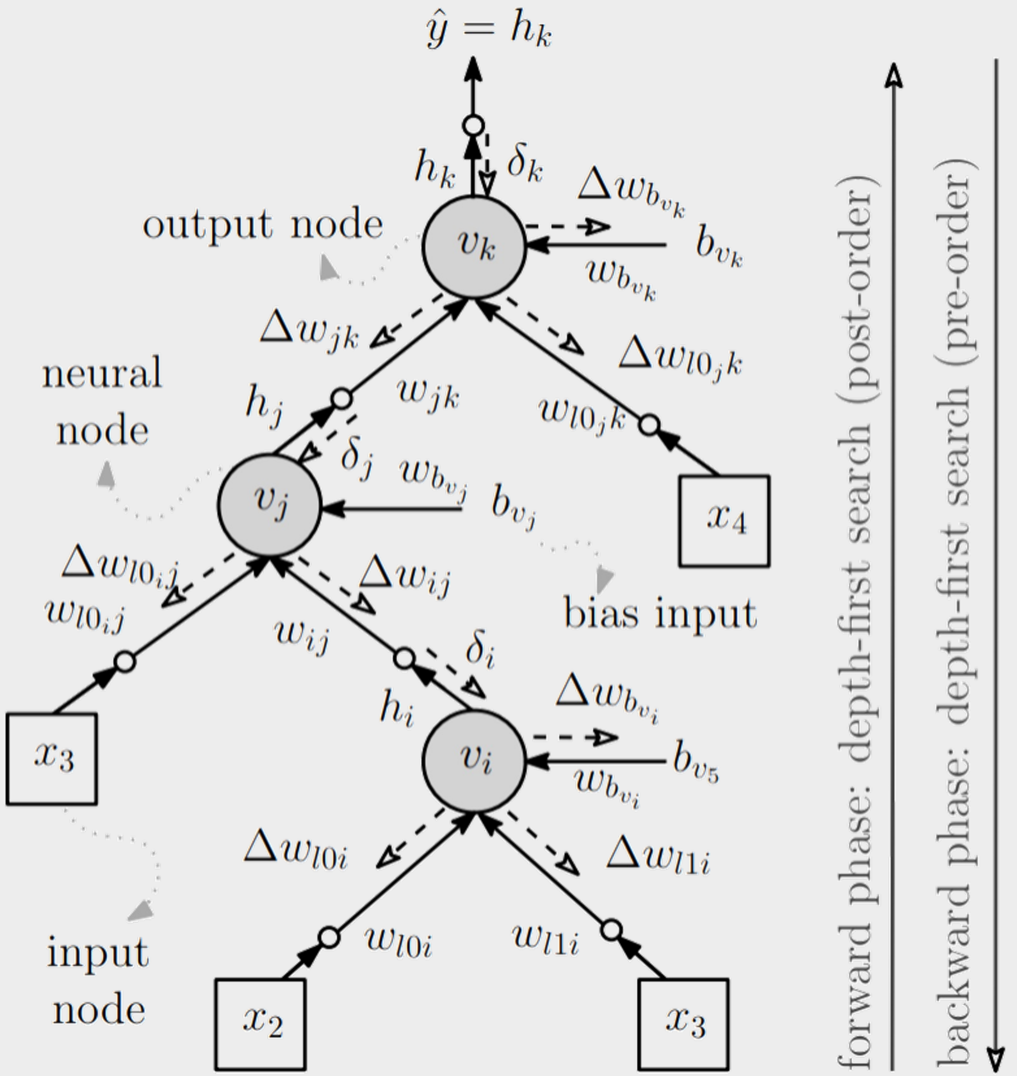
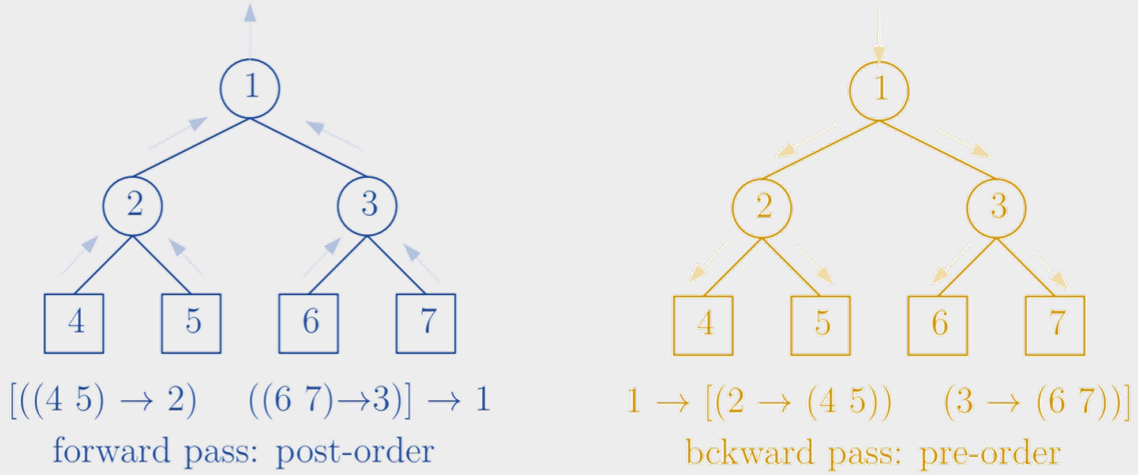


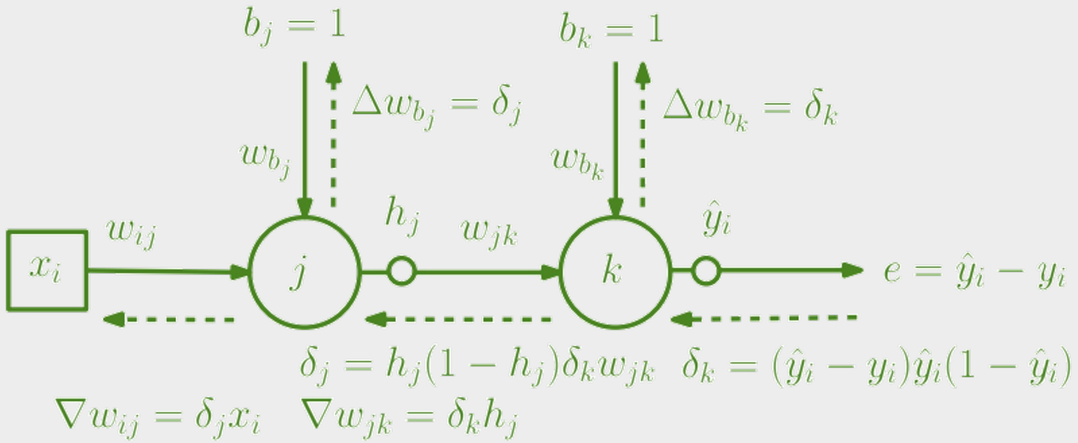
Fig A. Forward pass and gradient backpropagation



Backpropagation Neural Tree

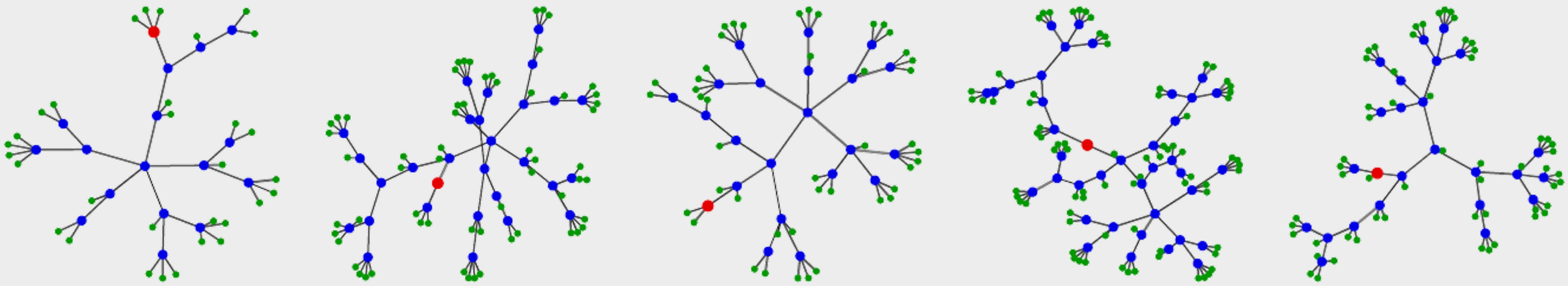






Backpropagation Neural Tree

Regression results



(a) baseball (.85, 48) (b) dee (.89, 89) (c) diabetese (.63, 67) (d) friedman (.95, 116) (e) mpg6 (.9, 82)

Algorithm	Bas	Dee	Dia	Frd	Mpg	Avg Acc	Avg Weights
BNeuralT	0.665	0.837	0.492	0.776	0.867	**0.727**	152
MLP	0.721	0.829	0.49	0.943	0.874	0.772	**1041**

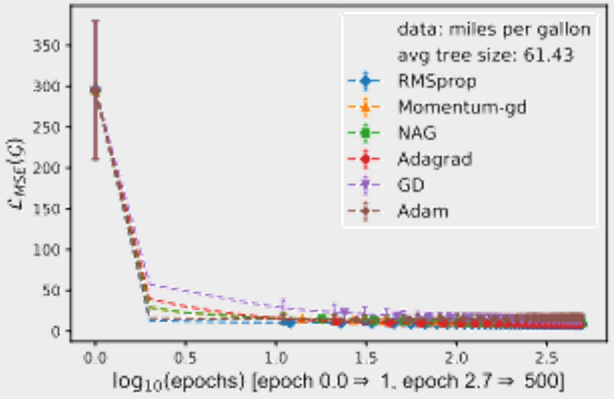
Backpropagation Neural Tree

Regression results

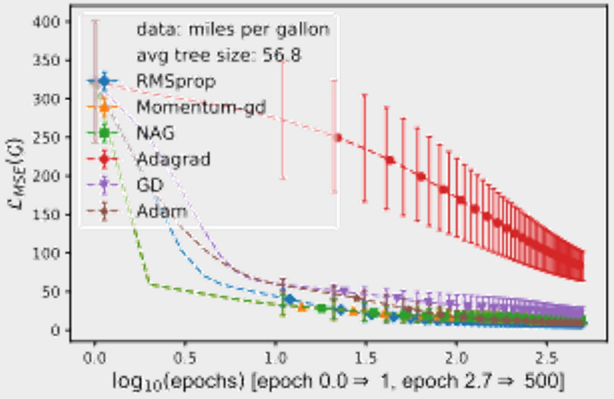
- BNeuralT used only 14.6% of MLP
- Accuracy differs only 5.8% lower than the best MLP result

Neural Tree vs Neural Networks

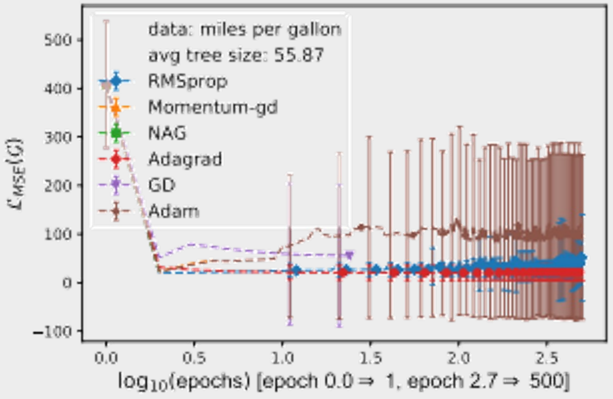
Regression Problems



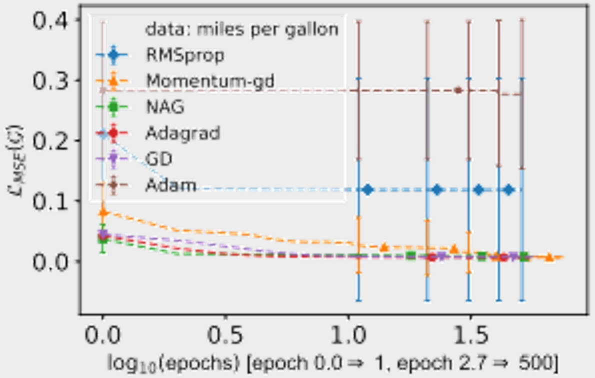
(g) BNeuralT: Sigmod, $\eta = 0.1$



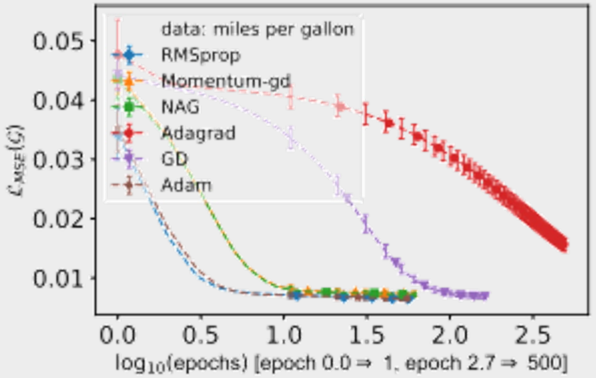
(h) BNeuralT: Sigmod, $\eta = default$



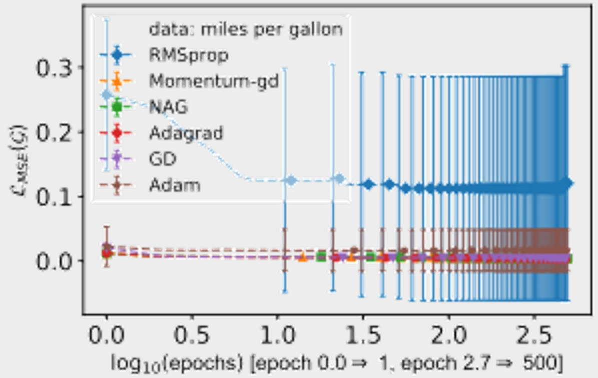
(i) BNeuralT: ReLU, $\eta = 0.1$



(j) MLP: Sigmod, $\eta = 0.1$



(k) MLP: Sigmod, $\eta = default$

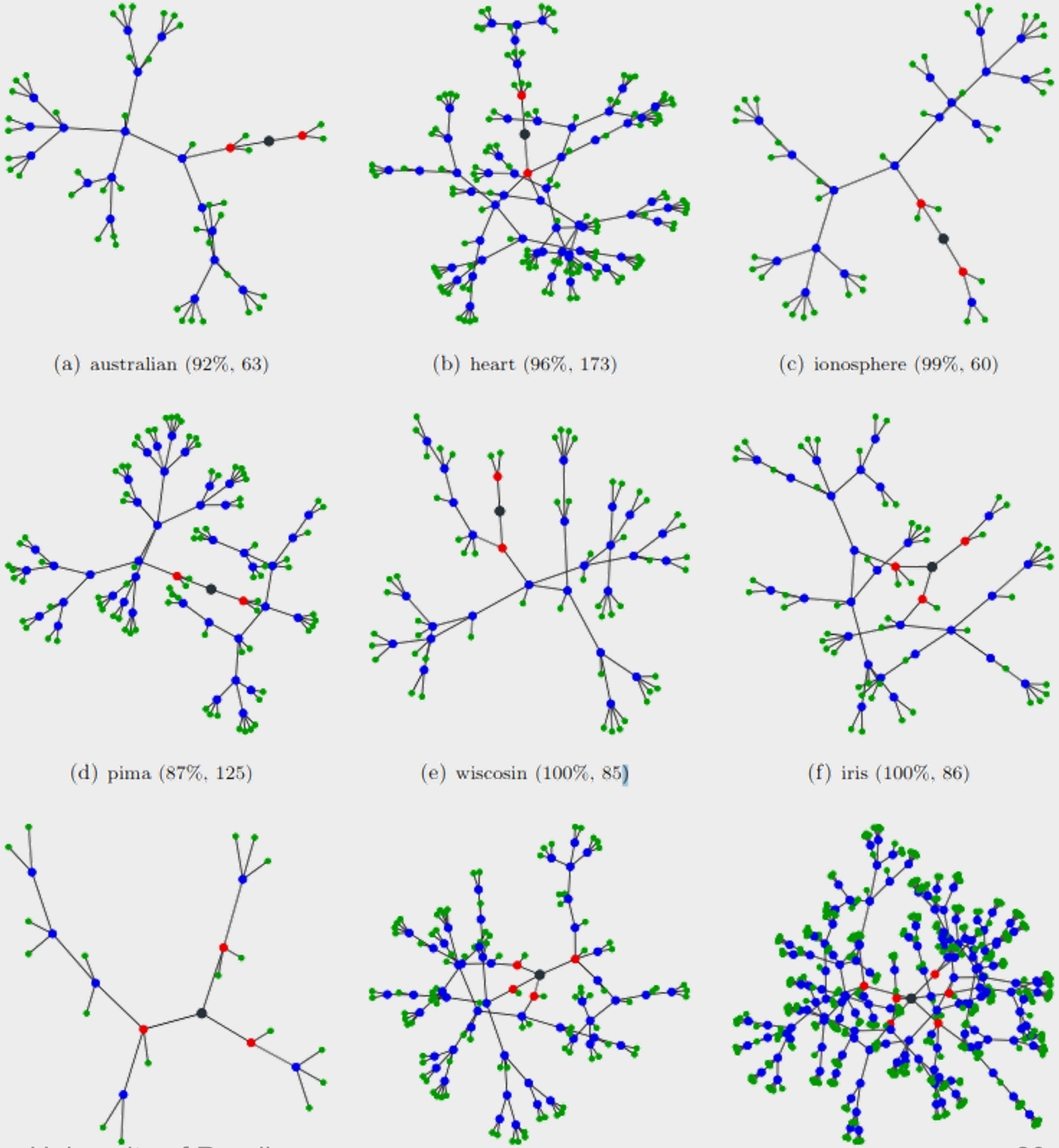


(l) MLP: ReLU, $\eta = 0.1$

Backpropagation Neural Tree

Classification results.

Data	BNeuralT	MLP
Aus	0.895	0.876
Hrt	0.897	0.833
Ion	0.952	0.882
Pma	0.822	0.774
Wis	0.986	0.984
Irs	0.992	0.972
Win	0.991	0.991
Vhl	0.75	0.826
Gls	0.732	0.635
Avg. Accuracy	**0.891**	**0.863**
Avg. Weights	**261**	1969



(a) australian (92%, 63)
(b) heart (96%, 173)
(c) ionosphere (99%, 60)
(d) pima (87%, 125)
(e) wiscosin (100%, 85)
(f) iris (100%, 86)

Ojha et al (2022), Neural Networks

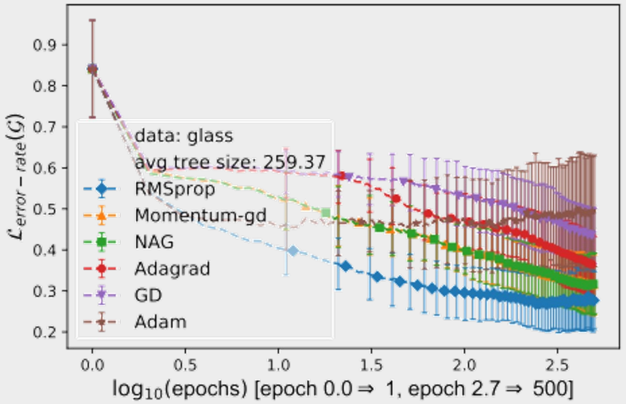
Backpropagation Neural Tree

Classification results

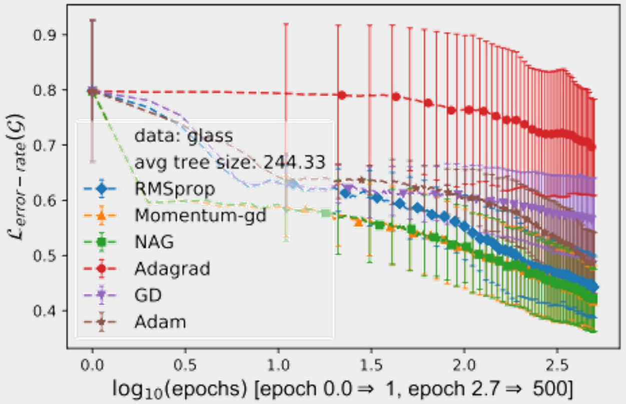
- BNeuralT used **only 13.25% parameters** of MLP
- Accuracy is **2.65% better than the best MLP** result

Neural Tree vs Neural Networks

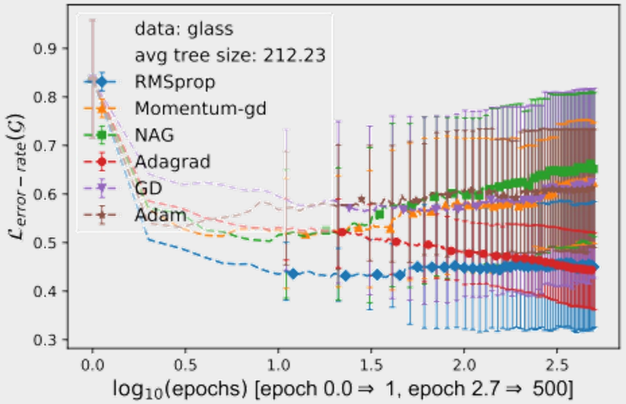
Classification Problems



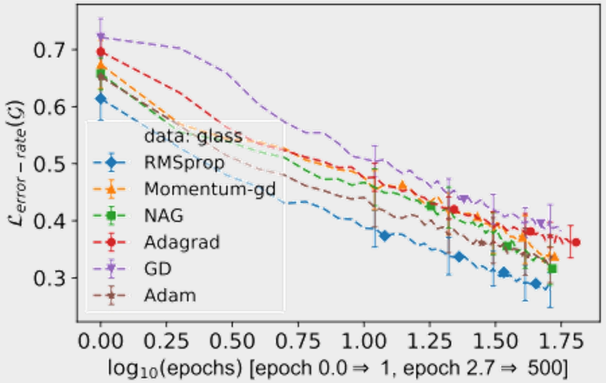
(a) BNeuralT: Sigmod, $\eta = 0.1$



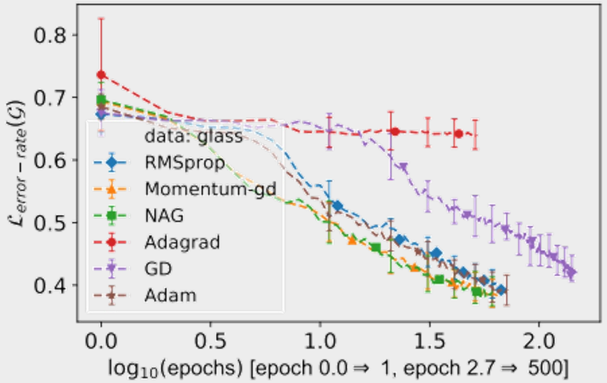
(b) BNeuralT: Sigmod, $\eta = default$



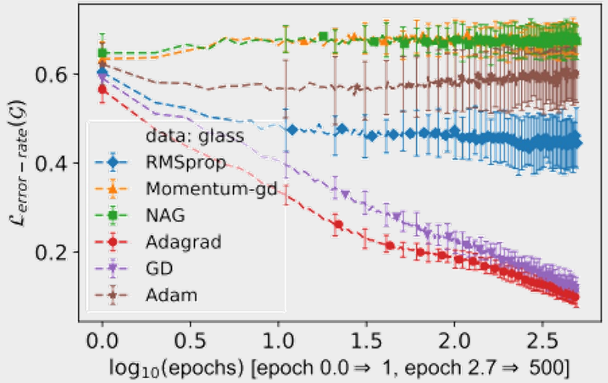
(c) BNeuralT: ReLU, $\eta = 0.1$



(d) MLP: Sigmod, $\eta = 0.1$

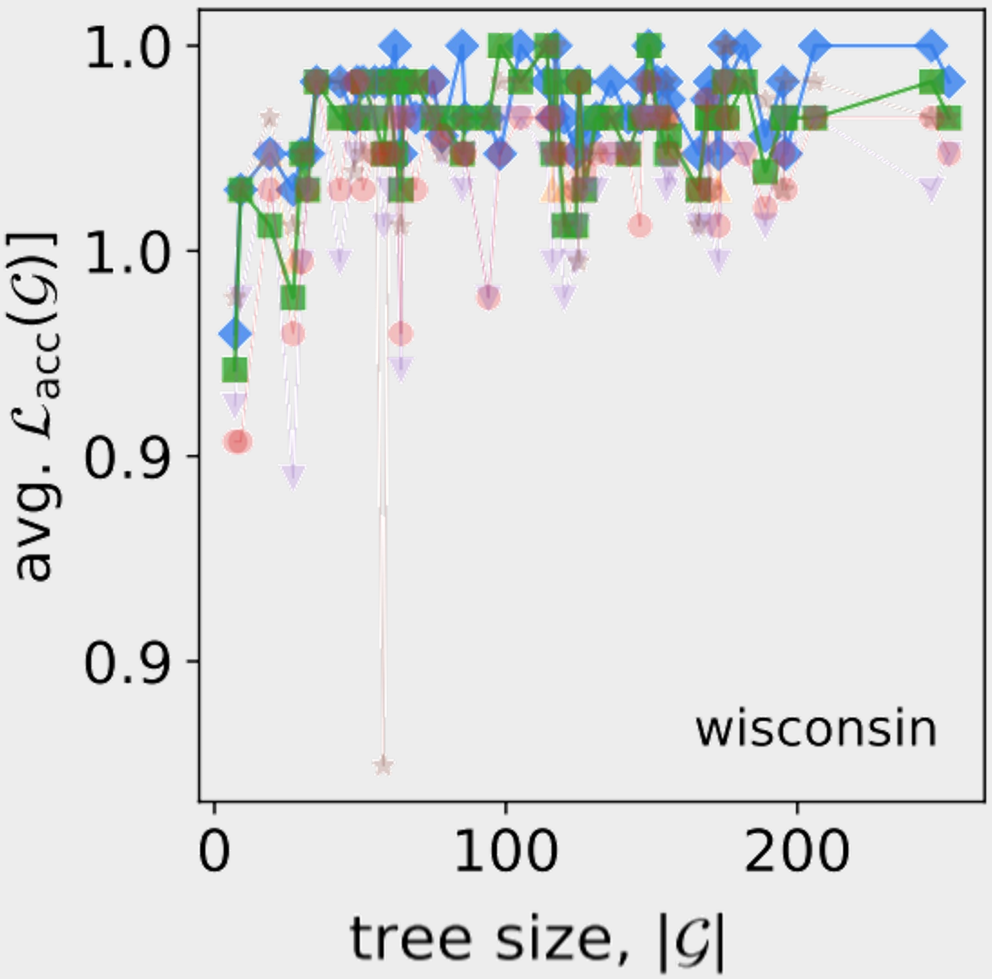


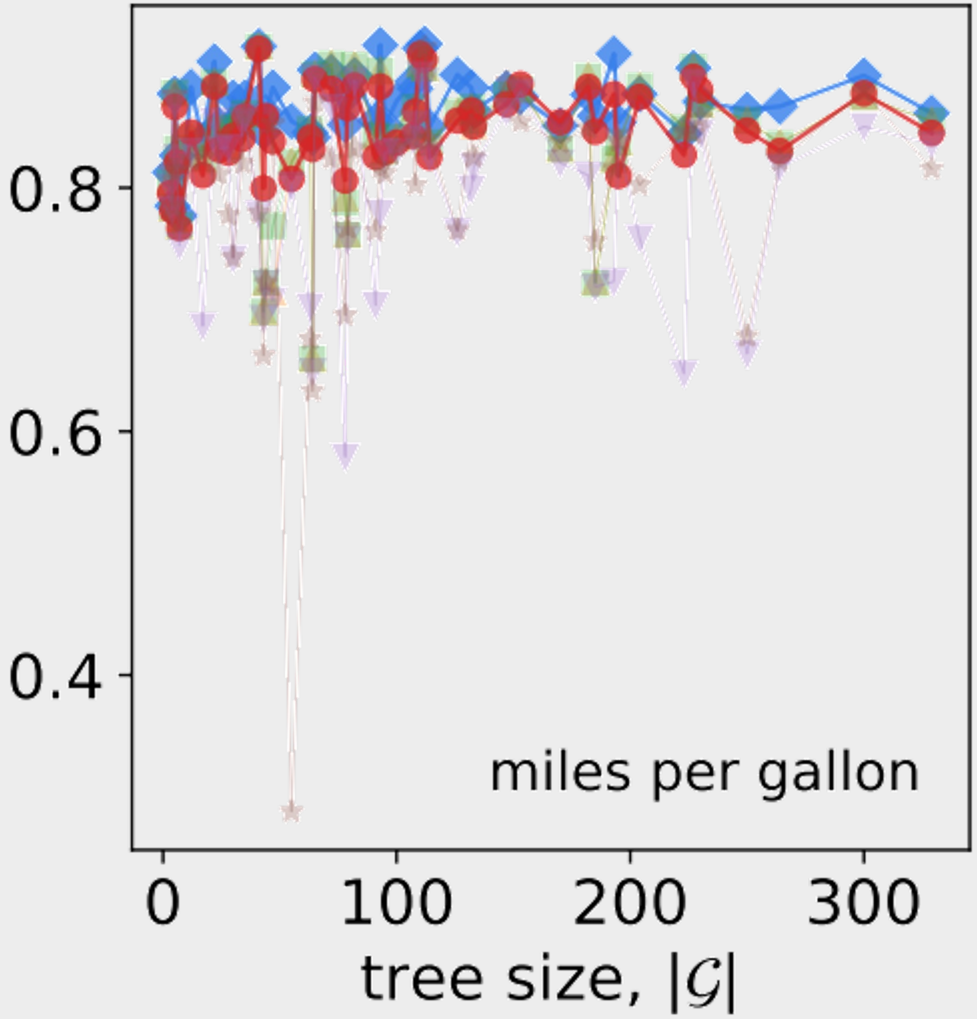
(e) MLP: Sigmod, $\eta = default$



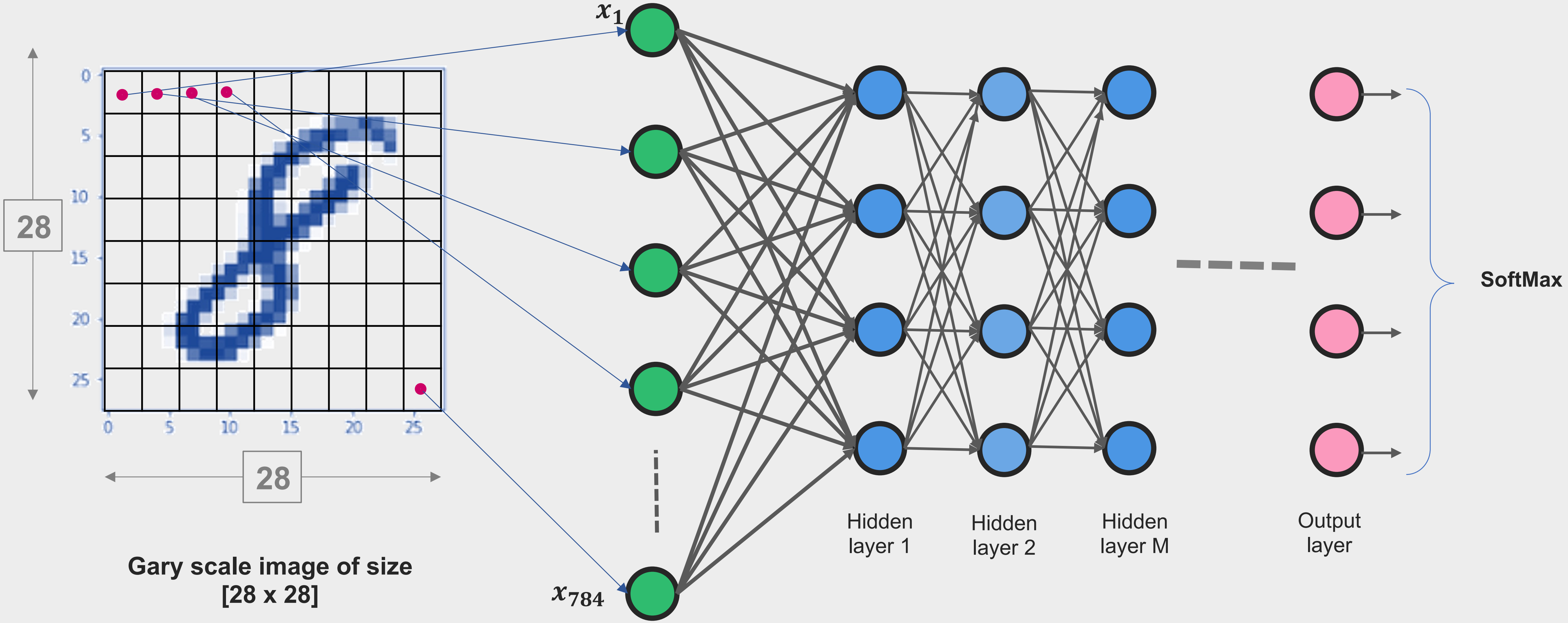
(f) MLP: ReLU, $\eta = 0.1$

Architectural Stochasticity





Deep Neural Networks



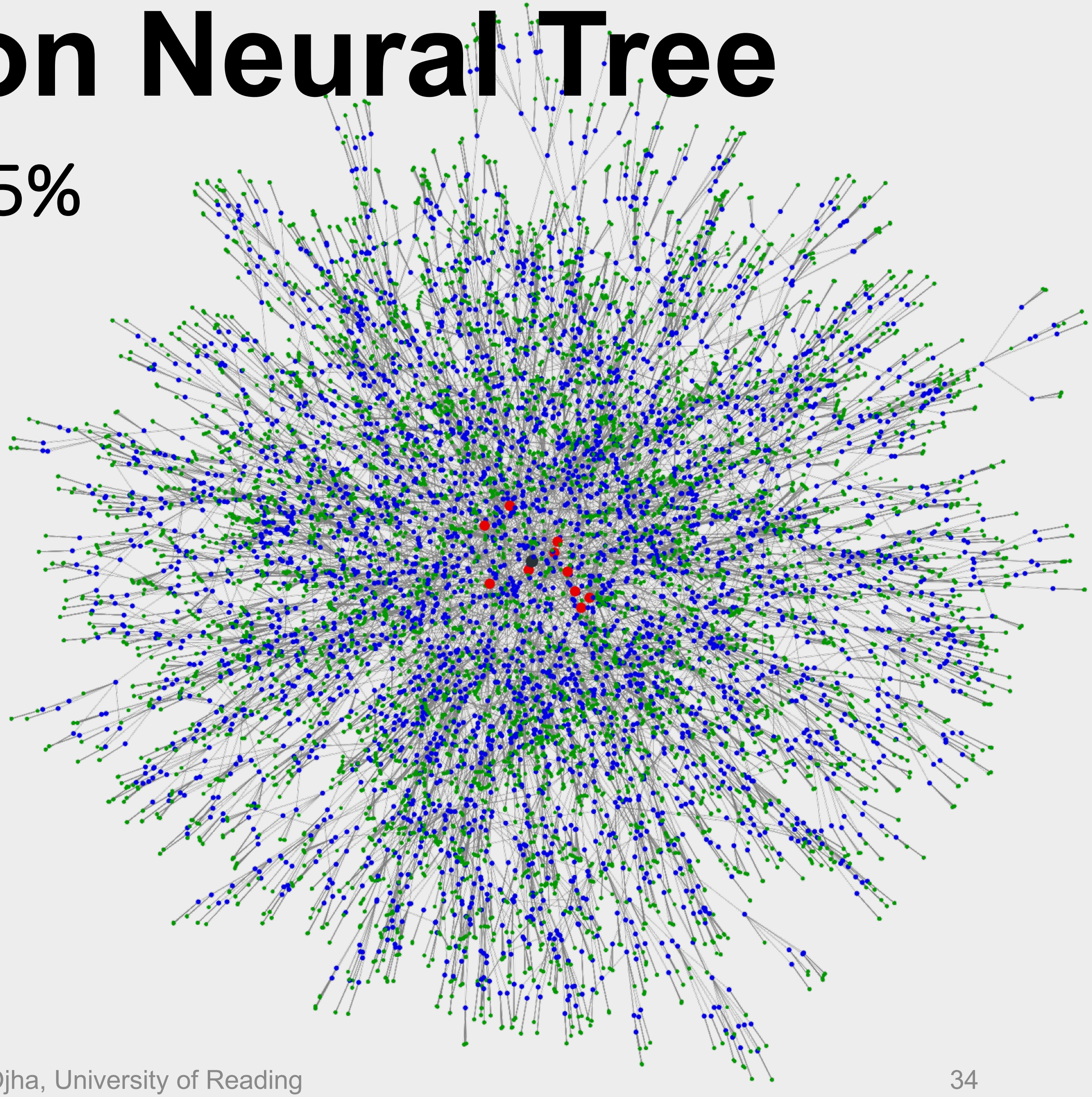
Gary scale image of size [28 x 28]

input layer

Backpropagation Neural Tree

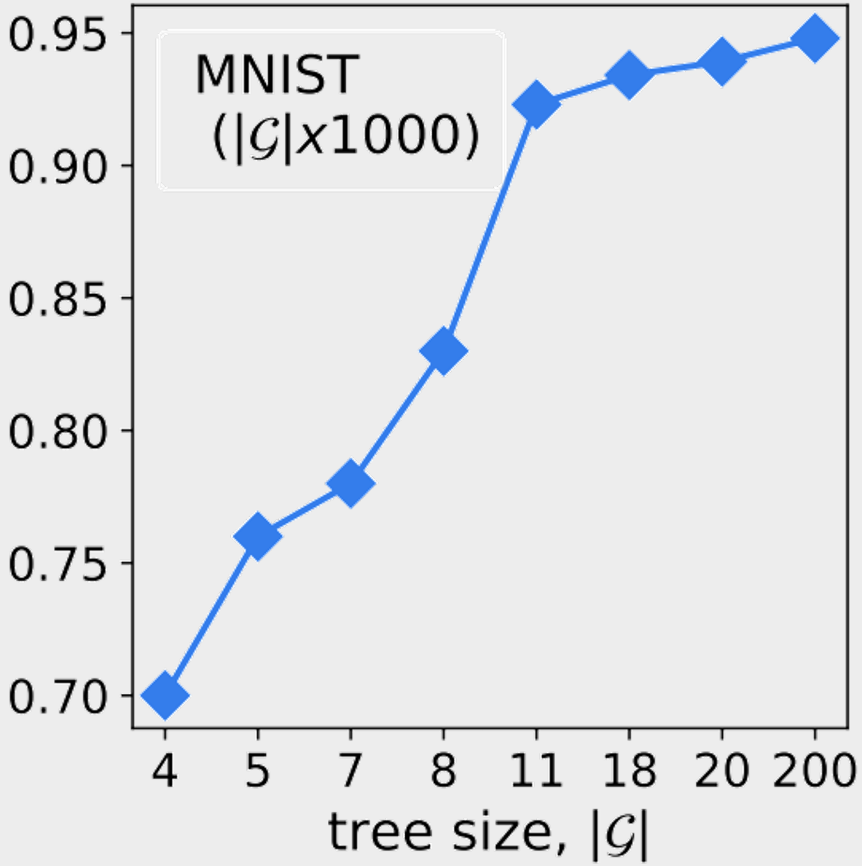
MNIST Model Accuracy ~95%

	Algorithms	Error(%)
BNeuralTs	BNeuralT-10K (pixels)	7.74
	BNeuralT-18K (pixels)	6.58
	BNeuralT-20K (pixels)	6.08
	BNeuralT-200K† (pixels)	**5.19**
Classification Trees	GUIDE (pixels, oblique split)	26.21
	OC1 (pixels, oblique split)	25.66
	GUIDE (pixels)	21.48
	CART-R (pixels)	11.97
	CART-P (pixels)	11.95
	C5.0 (pixels)	11.69
	TAO (pixels)	11.48
	TAO (pixels, oblique split)	5.26



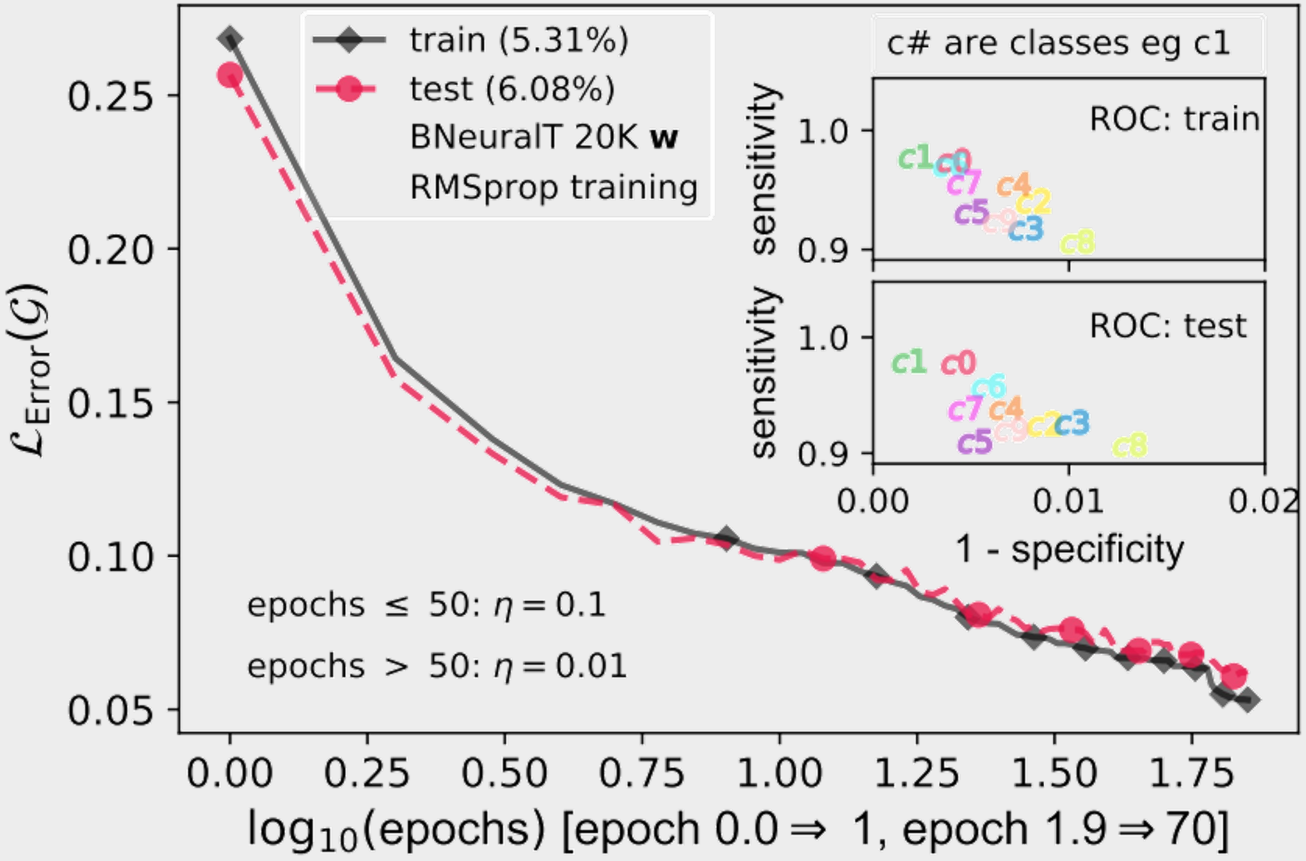
Model Size vs Accuracy

	Algorithms	Error(%)
BNeuralTs	BNeuralT-10K (pixels)	7.74
	BNeuralT-18K (pixels)	6.58
	BNeuralT-20K (pixels)	6.08
	BNeuralT-200K† (pixels)	**5.19**
Classification Trees	GUIDE (pixels, oblique split)	26.21
	OC1 (pixels, oblique split)	25.66
	GUIDE (pixels)	21.48
	CART-R (pixels)	11.97
	CART-P (pixels)	11.95
	C5.0 (pixels)	11.69
	TAO (pixels)	11.48
	TAO (pixels, oblique split)	5.26



Learnability of Different Classes

Competition between classes: TPR/Recall/Sensitivity vs FPR/(1 - Specificity)



Summary

stochastic gradient descent training of any a priori arbitrarily “thinned” network has the potential to solve machine learning tasks with an equivalent or better degree of accuracy than a fully connected symmetric and systematic neural network architecture.

References

- Ojha, V., & Nicosia, G. (2022). Backpropagation neural tree. Neural Networks, 149, 66-83.

- Ojha, V., & Nicosia, G. (2020). Multi-objective optimisation of multi-output neural trees. In 2020 IEEE Congress on Evolutionary Computation (CEC) (pp. 1-8). IEEE.

- Ojha, V. K., Abraham, A., & Snášel, V. (2017). Ensemble of heterogeneous flexible neural trees using multiobjective genetic programming. Applied Soft Computing, 52, 909-924.

- Ojha, V. K., Snášel, V., & Abraham, A. (2017). Multiobjective programming for type-2 hierarchical fuzzy inference trees. IEEE Transactions on Fuzzy Systems, 26(2), 915-936.