ICASSP 2021

TORONTO

Canada

June 6-11, 2021

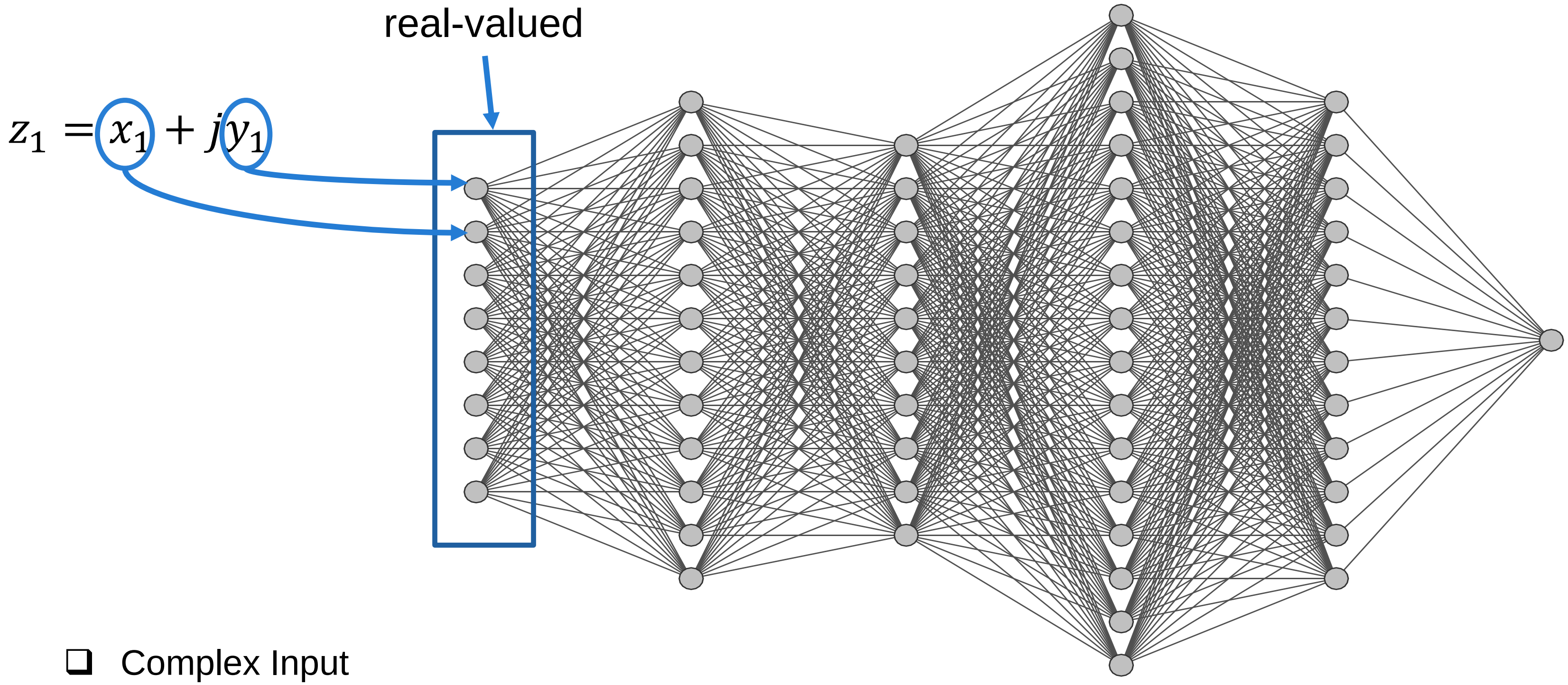
Metro Toronto Convention Centre

IEEE Signal Processing Society™

Complex-Valued Neural Network for Classification Perspectives: An Example on Non-Circular Data

J. Agustin BARRACHINA
Chengfang REN
Christèle Morisseau
Gilles Vieillard
Jean-Philippe Ovarlez

Mathematical Background



- ❑ Complex Input
- ❑ Complex Weights
- ❑ Complex activation functions
- ☒ Complex Learning Algorithm

$$\begin{cases} f(z) = g(x) + jh(y) & \text{(Type A)} \\ f(z) = g(|z|)e^{j\arg(z)} & \text{(Type B)} \end{cases}$$

[2] Kuroe et al. *"On Activation Functions for Complex-Valued Neural Networks"* 2003

Mathematical Background

Liouville's theorem:
"Given $f(z)$ analytic (differentiable) at all $z \in C$ and bounded, then $f(z)$ is a constant function"

Wirtinger Calculus

$$\frac{\delta f}{\delta z} \triangleq \frac{1}{2}\left(\frac{\delta f}{\delta x} - j\frac{\delta f}{\delta y}\right) \qquad \frac{\delta f}{\delta \bar{z}} \triangleq \frac{1}{2}\left(\frac{\delta f}{\delta x} + j\frac{\delta f}{\delta y}\right)$$

$$\nabla_Z(\cdot) = \begin{bmatrix} {}^{\delta}/_{\delta x_1} + j\,{}^{\delta}/_{\delta y_1} \\ \vdots \\ {}^{\delta}/_{\delta x_n} + j\,{}^{\delta}/_{\delta y_n} \end{bmatrix} \qquad \nabla_z f \triangleq 2\frac{\delta f}{\delta \bar{z}} = \frac{\delta f}{\delta x} + j\frac{\delta f}{\delta y}$$

[3] Robert F. H. Fischer *"Precoding and Signal Shaping for Digital Transmission"* 2002

Motivation

APPLICATIONS OF COMPLEX-VALUED NEURAL NETWORKS

Applications	**Corresponding Publications**
Radio Frequency Signal Processing in Wireless Communications	[6], [8], [11], [32], [33], [35], [36], [52], [53], [55], [56], [65]–[67], [72], [89], [91], [109], [123]–[128], [133], [149]–[152], [154], [155]
Image Processing and Computer Vision	[19], [31], [34], [37], [39], [40], [54], [62], [69], [73], [75]–[77], [82], [85], [92], [98], [99], [101]–[103], [119], [130], [141], [142], [156]–[159]
Audio Signal Processing and Analysis	[26], [48], [49], [58], [79], [130], [136]
Radar / Sonar Signal Processing	[74], [110], [139], [153], [160], [161]
Cryptography	[162]
Time Series Prediction	[103], [139]
Associative Memory	[105], [116]
Wind Prediction	[30], [43], [148]
Robotics	[38]
Traffic Signal Control (robotics)	[46], [60]
Spam Detection	[59]
Precision Agriculture (soil moisture prediction)	[82]

[1] Bassey et al. *“A Survey of Complex-Valued Neural Networks”* 2021.

[2] Mönning et al. *“Evaluation of Complex-Valued Neural Networks on Real-Valued Classification Tasks”* 2018.

[3] El-Darymli et al. *“On circularity/noncircularity in single-channel synthetic aperture radar imagery”* 2014.
[4] Vasile et al. *“Circularity of complex stochastic models in polsar and multi-pass insar images”* 2012.
[5] Barbaresco et al. “Noncircularity exploitation in signal processing overview and application to radar” 2008.
[6] Wu et al. “Noncircularity parameters and their potential applications in UHR MMW SAR data sets” 2016.

[7] Hirose “Complex-Valued Neural Networks” 2012.
[8] Hirose “Complex-Valued Neural Networks: Advances and applications” 2013.

?

ONERA
THE FRENCH AEROSPACE LAB

Mathematical Background: Circularity

Complex random variable $\mathrm{Z} = X + jY$ is *circular* if
Z has the same distribution as $e^{j\phi}\mathrm{Z}$

$$\varrho_Z = \frac{\tau_Z}{\sigma_Z}\begin{cases} = 0 & \rightarrow z \; is \; circular \\ \neq 0 & \rightarrow z \; not \; circular \end{cases}$$

- $\tau_Z \triangleq \mathrm{E}[(Z - \mathrm{E}[Z])^2] = \sigma_X^2 - \sigma_Y^2 + 2j\sigma_{XY}$
- ~~$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2$~~

Two sources of non-circularity [1]:

1. Unequal variances
2. Correlation $\rho = {}^{E[xy]}/_{\sqrt{E[x^2]E[y^2]}}$

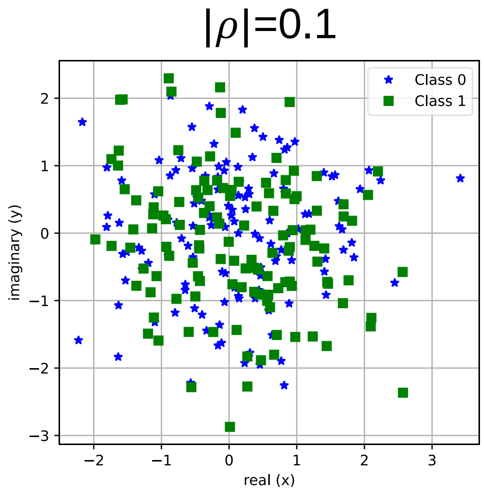
[4] Ollila "On the Circularity of a Complex Random Variable" 2008

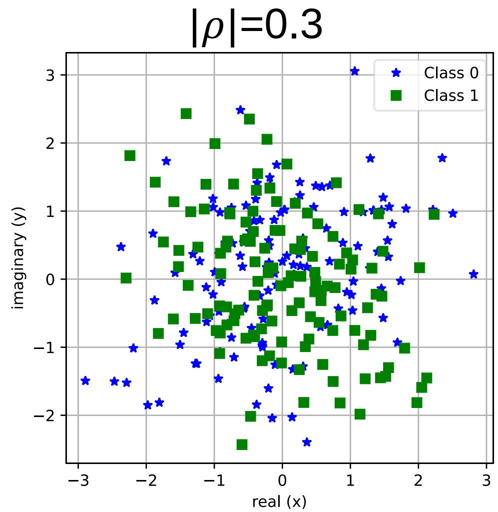
Mathematical Background: Circularity

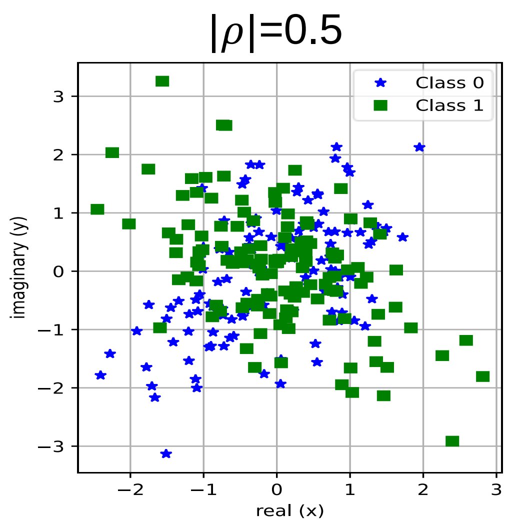
1 vector of class 0: $|\rho|$

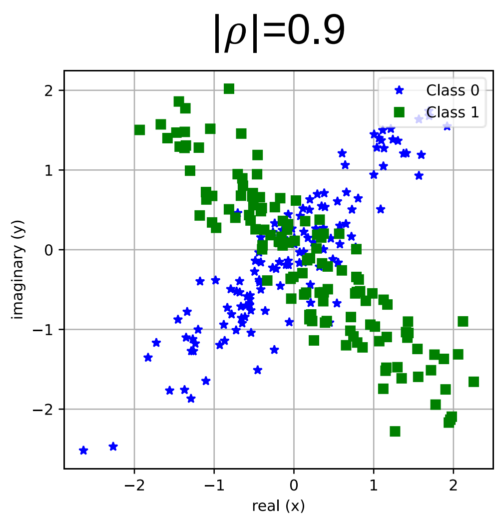
1 vector of class 1: $-|\rho|$

*Note:
Each point on the graph corresponds to one component of the input vector









Model Architecture

Complex-Valued Multi-Layer Perceptron

Model:

- Loss: Categorical cross-entropy
- Weight initialization: Glorot uniform
- SGD (Stochastic Gradient Descent)
  - Learning rate 0.1
  - Wirtinger Derivative

	CVNN	RVNN
Input Size	128	256
Hidden Layer Size	64	128
Activation Function	ReLU Type A [2]	ReLU
Dropout	50%	50%
Output Size	2	2
Output Activation	Softmax over absolute value	Softmax

Dataset:

- Input vector size 128
- 8000 training vectors / class
- 2000 validation vectors / class

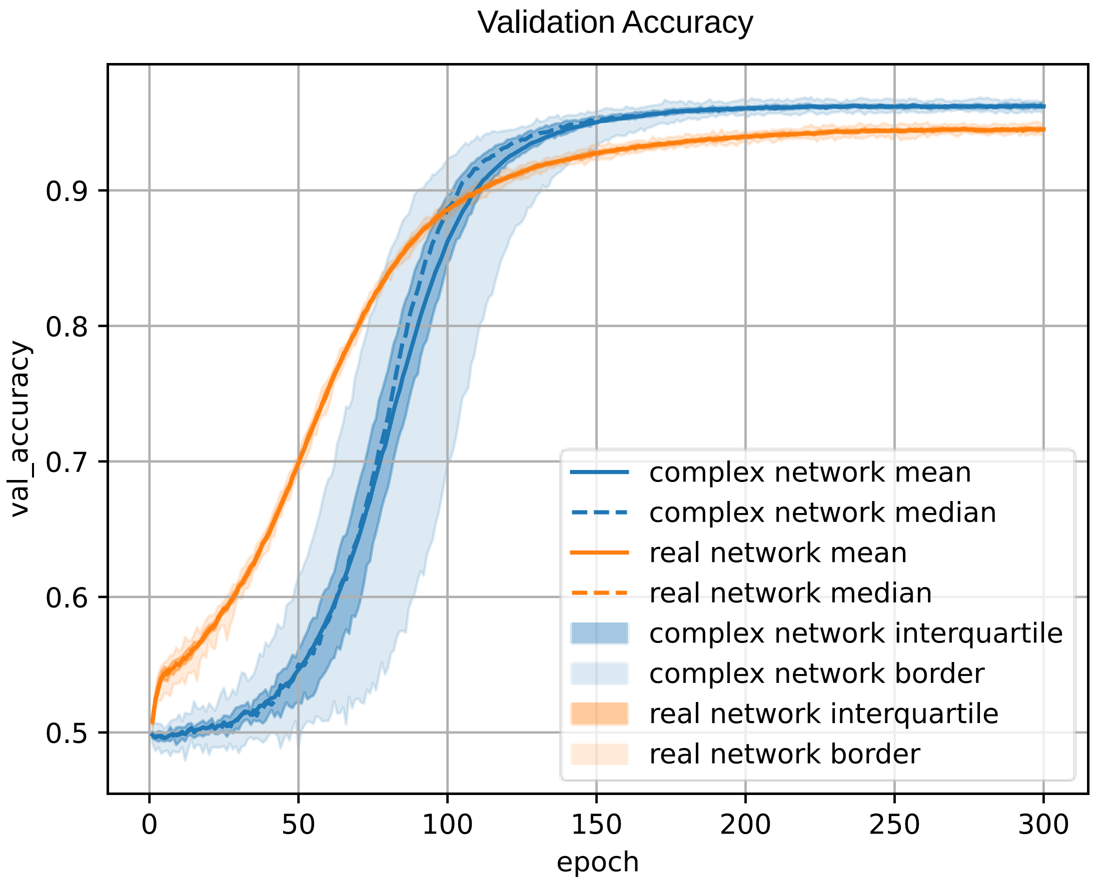
Simulation:

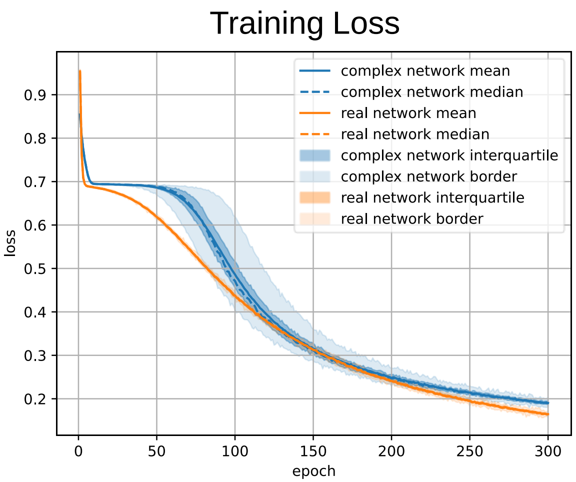
- 30 trials each model
- 300 epochs
- Batch size 100

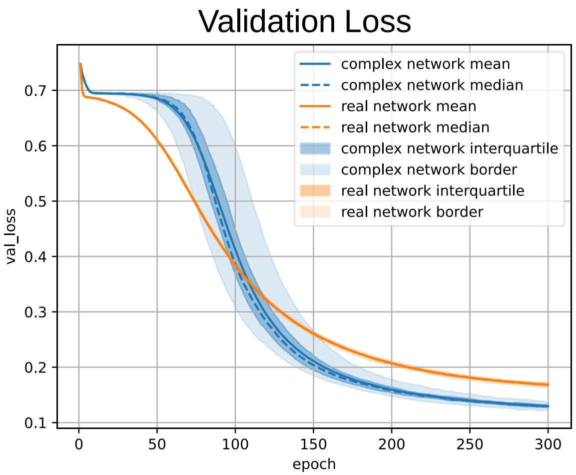
[2] Kuroe et al. *“On Activation Functions for Complex-Valued Neural Networks”* 2003

Experimental Results

$|\rho| = 0.3$

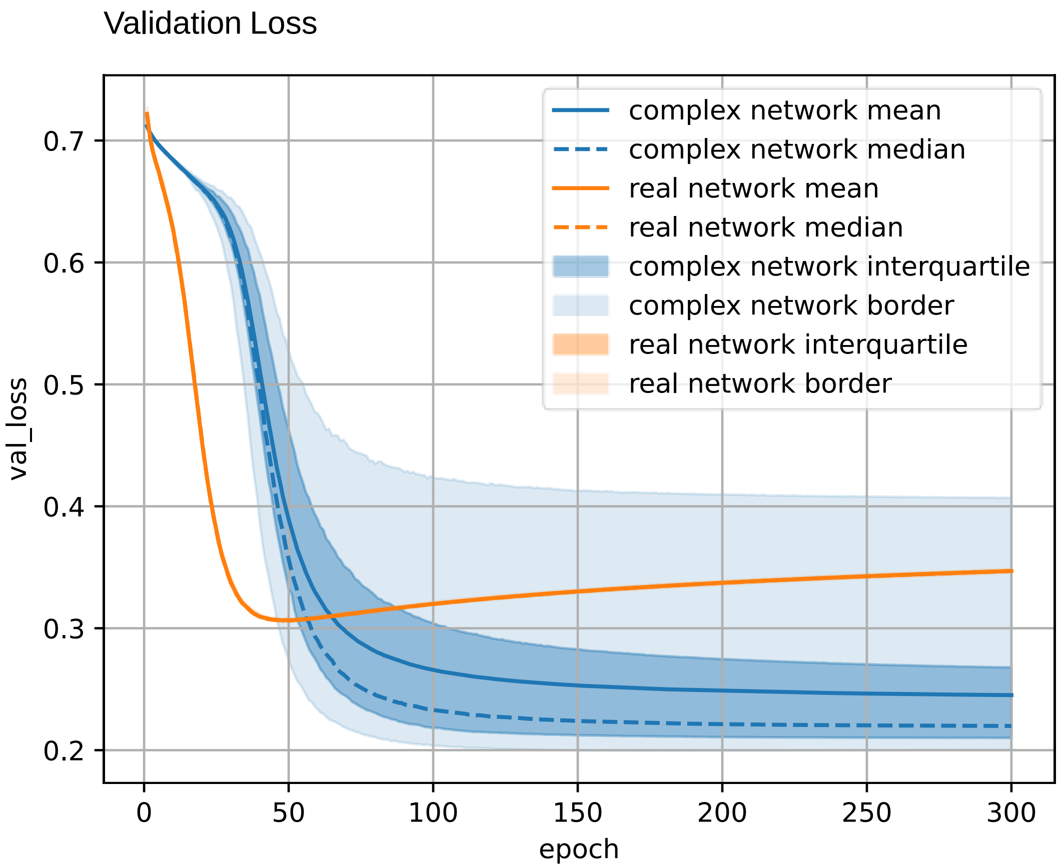






ONERA THE FRENCH AEROSPACE LAB

Experimental Results: Without Dropout



Experimental Results: Different Datasets

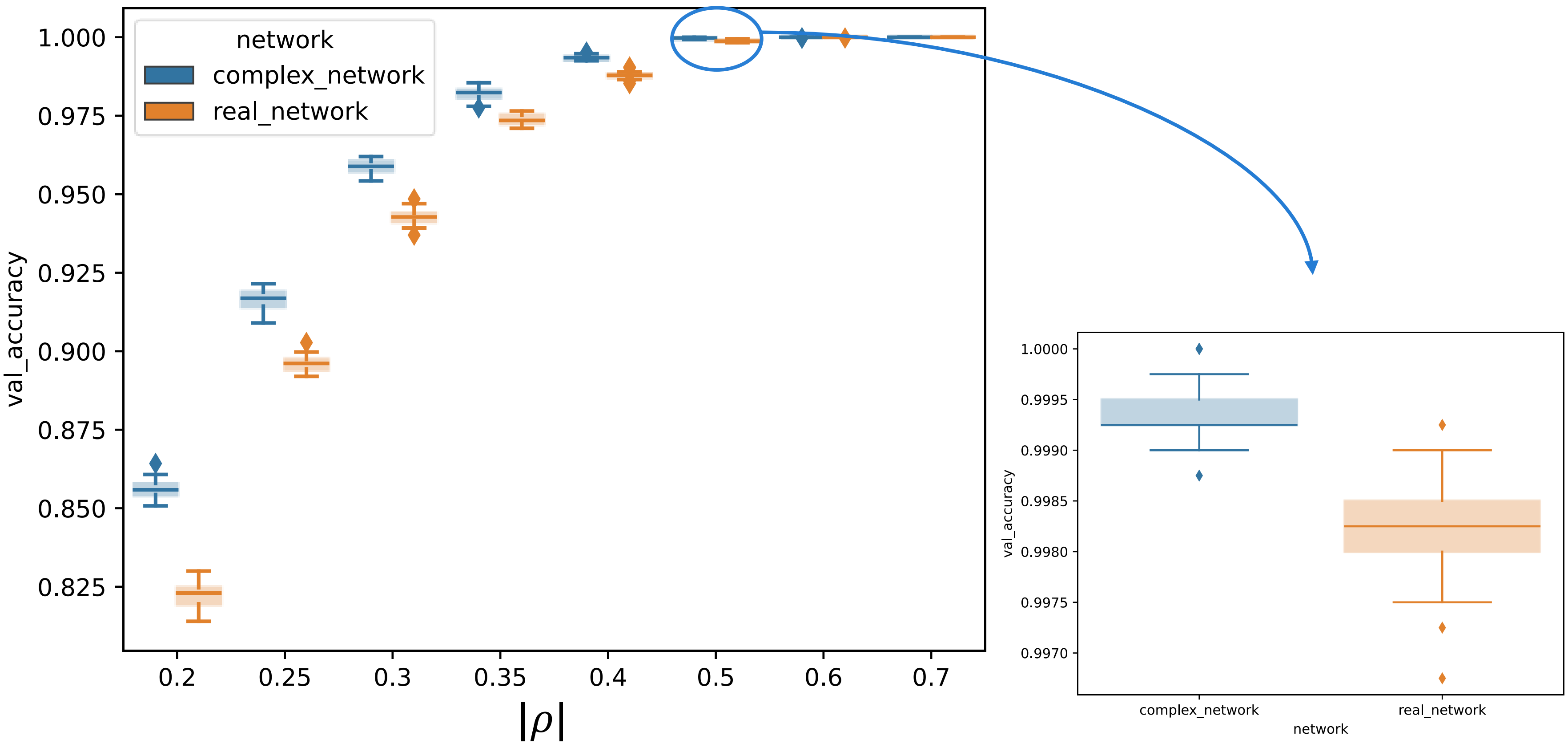
	Data A		Data B		Data C	
Class	1	2	1	2	1	2
ρ	0.3	-0.3	0	0	0.3	-0.3
σ_X^2	1	1	1	2	1	2
σ_Y^2	1	1	2	1	2	1
ϱ_Z	$j0.3$	$-j0.3$	$-\frac{1}{3}$	$\frac{1}{3}$	$\frac{j-0.6}{3}$	$\frac{0.6-j}{3}$

Table 1: Dataset Characteristics

		Data A		Data B		Data C	
		CVNN	RVNN	CVNN	RVNN	CVNN	RVNN
1HL	median	96.16 ± 0.11	94.48 ± 0.06	97.39 ± 0.07	96.65 ± 0.06	99.67 ± 0.04	99.48 ± 0.03
	mean	96.20 ± 0.04	94.51 ± 0.04	99.67 ± 0.02	96.66 ± 0.04	96.66 ± 0.01	99.47 ± 0.02
	IQR	$96.06 - 96.43$	$94.38 - 94.59$	$97.31 - 97.54$	$96.57 - 96.78$	$99.61 - 99.73$	$99.43 - 99.52$
	full range	$95.65 - 96.60$	$94.02 - 95.03$	$97.03 - 97.80$	$96.25 - 97.07$	$99.50 - 99.77$	$99.27 - 99.67$
2HL	median	97.83 ± 0.08	95.82 ± 0.08	98.89 ± 0.05	97.83 ± 0.05	99.90 ± 0.02	99.87 ± 0.01
	mean	97.81 ± 0.04	95.86 ± 0.04	98.88 ± 0.02	97.82 ± 0.03	99.90 ± 0.01	99.86 ± 0.01
	IQR	$97.70 - 97.97$	$95.71 - 95.97$	$98.77 - 98.94$	$97.75 - 97.90$	$99.87 - 99.92$	$99.84 - 99.87$
	full range	$97.35 - 98.22$	$95.53 - 96.30$	$98.65 - 99.05$	$97.62 - 98.08$	$99.85 - 99.98$	$99.77 - 99.92$

Table 2: Test accuracy results (%)

Experimental Results: Correlation Coefficient swipe



Conclusions

- Almost 100 cases tested
  - Different source of Non-Circularity
  - Without Dropout
  - Number of layers
  - Amplitude-Phase
  - Size of the hidden layers
  - Activation function
  - Learning rate
- CVNNs generalize better
  - They tend to take more epochs to reach stability
- In general, cases where RVNN outperformed CVNN
  - Under 60% accuracy

Thank you!