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

J. Agustin BARRACHINA*†‡, Chengfang REN†, Christele MORISSEAU*, Gilles VIEILLARD*, Jean-Philippe OVARLEZ*†

*DEMR, ONERA, Université Paris-Saclay, F-91123, Palaiseau, France

ONERA THE FRENCH AEROSPACE LAB

†SONDRA, CentraleSupelec, Université Paris-Saclay Gif-sur-Yvette, France

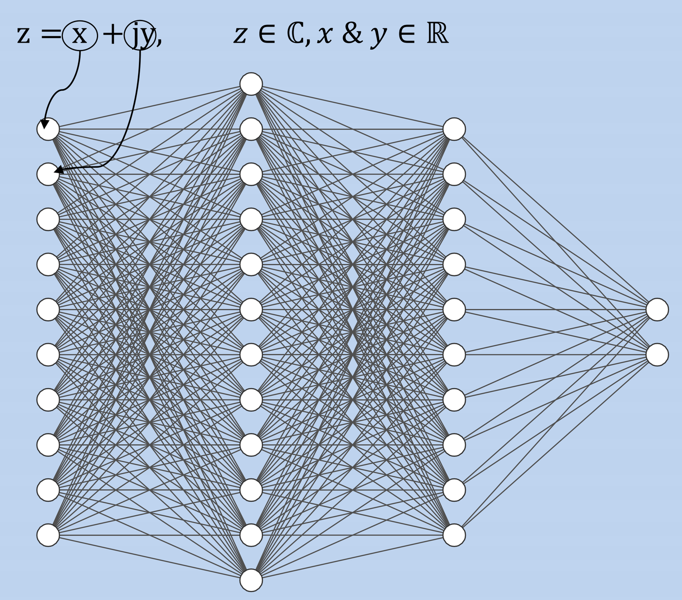
SONDRA CentraleSupélec

‡The author would like to thank the DGA for funding DGA

Abstract – This paper shows the benefits of using Complex-Valued Neural Networks (CVNN) on classification tasks for non-circular complex-valued datasets. Motivated by radar and especially Synthetic Aperture Radar (SAR) applications, we propose a statistical analysis of fully connected feed-forward neural networks performance in the cases where real and imaginary parts of the data are correlated through the non-circular property.
In this context, comparisons between CVNNs and their real-valued equivalent models are conducted, showing that CVNNs provide better performance for multiple types of non-circularity. Notably, CVNNs statistically perform less overfitting and higher accuracy than its equivalent RVNN.

Context

The vast majority of neural networks architectures is based on representation of real valued features.



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.

Can complex-valued neural networks exploit phase information to achieve better results than real-valued neural networks?

Mathematical Background

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

Liouville theorem forces the activation functions to be a constant for the gradient to exist (needed for backpropagation). This is of course unacceptable and therefore, a new definition of the gradient, with the help of *Wirtinger calculus*, is created to solve this problem.

Wirtinger Calculus: $\frac{\partial f}{\partial z} = \frac{1}{2}\left(\frac{\partial f}{\partial x} - j\frac{\partial f}{\partial y}\right)$

Gradient definition: $\nabla_z f = 2\frac{\partial f}{\partial \bar{z}}$ for $f: C \to R$

Chain Rule: $\frac{\partial f \circ g}{\partial z} = \frac{\partial f}{\partial r}\frac{\partial r}{\partial z} + \frac{\partial f}{\partial s}\frac{\partial s}{\partial z}$

$$f: \mathbb{C} \to \mathbb{R};\ g(z) = r(z) + js(z);\ r, s: \mathbb{C} \to \mathbb{R}, z \in \mathbb{C}$$

Experimental Setup

Model Architecture

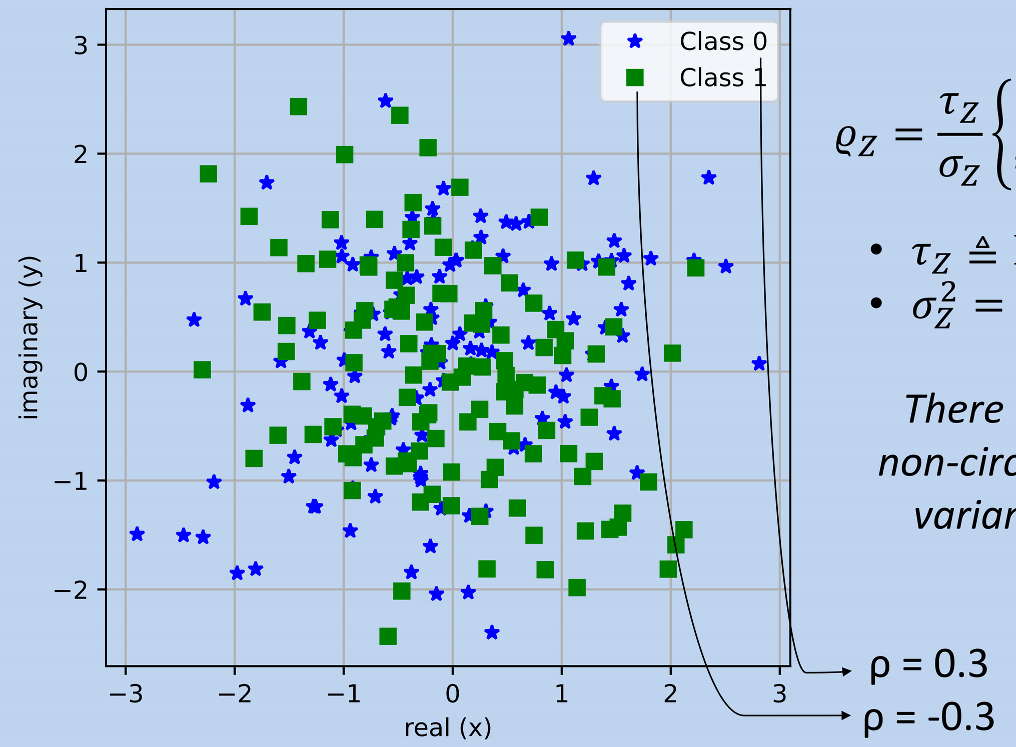
	CV-MLP	RV-MLP
Input Size	128	256
Hidden Layer 1	64	128
Activation function	ReLU type A [6]: $ReLU(\mathbb{R}e\{z\}) + jReLU(\mathbb{I}m\{z\})$	ReLU: $ReLU(x)$
Output size	2	2
Output activation	Softmax to the absolute value	Softmax

- Loss function:
  - Categorical cross-entropy
- Stochastic Gradient Descent
  - Learning rate: 0.1
- Weight initialization:
  - Glorot Uniform [7]
  - Bias initialization: Zeros
- 300 epochs
  - Batch Size: 100

Dataset

Akira Hirose has mentioned the importance of circularity[1] for CVNN in [1] section 1.4.2 and [3] section 3.2.2

Example of two input vectors



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

$$\varrho_Z = \frac{\tau_Z}{\sigma_Z}\begin{cases} = 0 \to z \text{ is circular} \\ \neq 0 \to z \text{ not circular}\end{cases}$$

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

There are two possible sources for non-circularity: x and y have unequal variances, and/or are x and y are correlated [5]

$$\rho = \frac{\sigma_{xy}}{\sigma_x\sigma_y} = \frac{E[xy]}{\sqrt{E[x^2]E[y^2]}}$$

References

[1] A. Hirose, *Complex-valued neural networks: Advances and applications*, vol. 18. John Wiley & Sons, 2013.

[2] A. Hirose and S. Yoshida, "Generalization Characteristics of Complex-Valued Feedforward Neural Networks in Relation to Signal Coherence," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 23, no. 4, pp. 541–551, Apr. 2012, doi: 10.1109/TNNLS.2012.2183613.

[3] A. Hirose, *Complex-Valued Neural Networks*. Springer Science & Business Media, 2012.

[4] R. Hansch and O. Hellwich, "Classification of Polarimetric SAR Data by Complex-Valued Neural Networks" p. 5, May 2009.

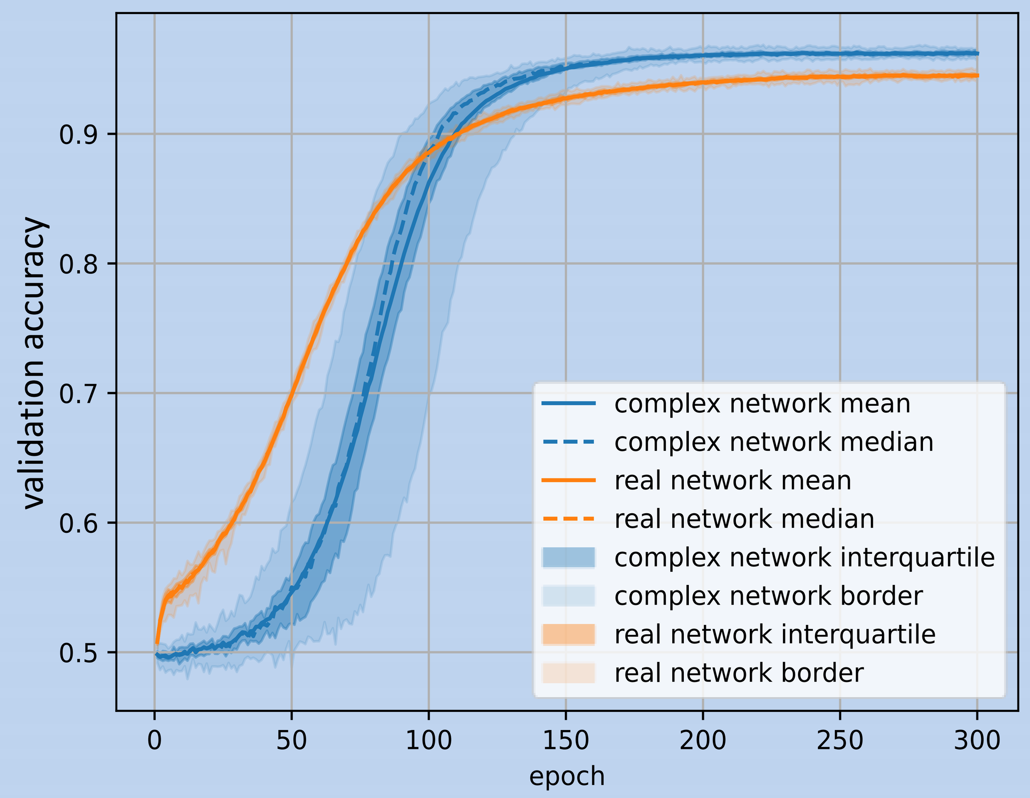
[5] E. Ollila, "On the Circularity of a Complex Random Variable," *IEEE Signal Process. Lett.*, vol. 15, pp. 841–844, 2008, doi: 10.1109/LSP.2008.2005050.

[6] Y. Kuroe, M. Yoshid, and T. Mori, "On Activation Functions for Complex-Valued Neural Networks — Existence of Energy Functions —," in *Artificial Neural Networks and Neural Information Processing — ICANN/ICONIP 2003*, Berlin, Heidelberg, 2003, pp. 985–992, doi: 10.1007/3-540-44989-2_117.

[7] X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010, pp. 249–256.

Results

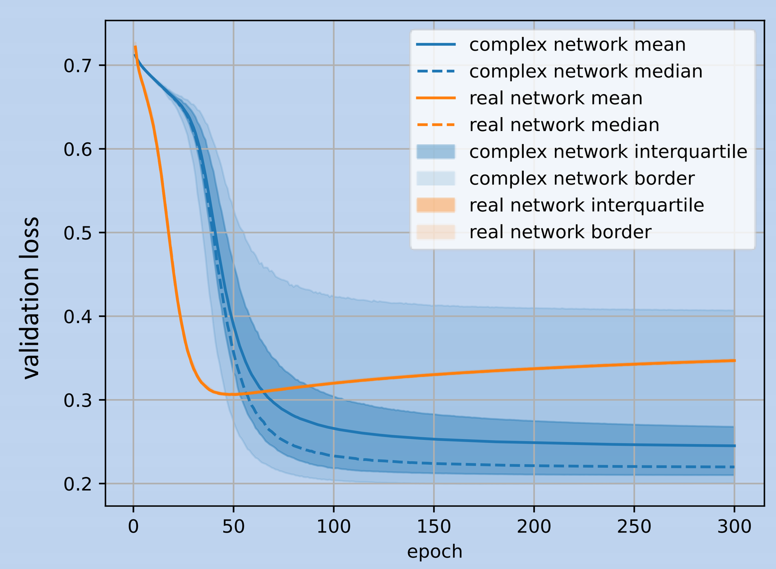
Validation accuracy with dropout



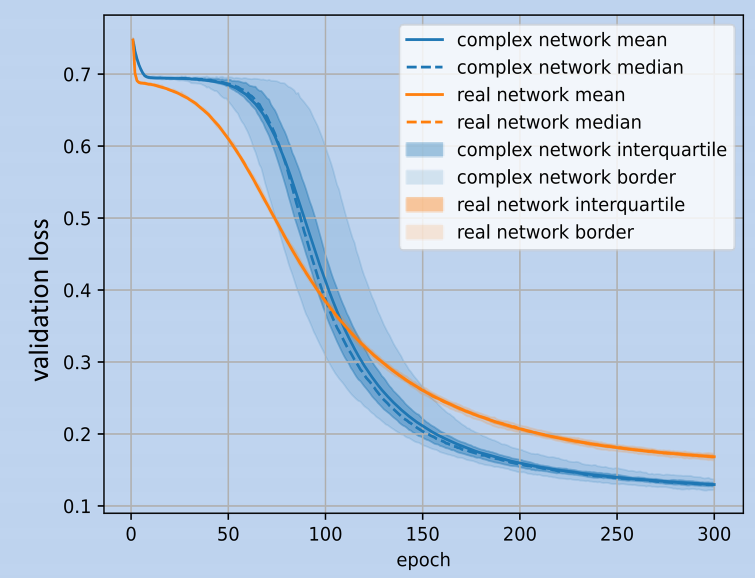
- 30 train trials for each model (CV-MLP & RV-MLP) with the same dataset.
- 20000 examples (10000 for each class)
  - 80% train set
  - 20% test set
- Input size of 128 independent non-circular random variables.
- RV-MLP gets real and imaginary part separately.

Model	Mean (%)	Median (%)	Q1 (25%)	Q2 (75%)	Min (%)	Max (%)
CV-MLP	96.20±0.04	96.16±0.11	96.06	96.43	95.65	96.60
RV-MLP	94.51±0.04	94.48±0.06	94.38	94.59	94.02	95.03

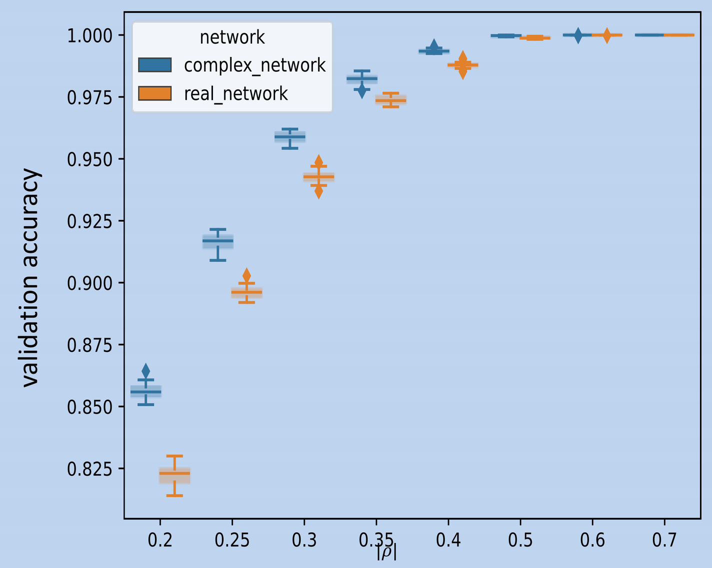
Validation loss without dropout



Validation loss with dropout



Without Dropout:
- CVNN presented less overfitting
- CVNN presented more outliers



Conclusions

- Skewed data (high difference between mean and median)
- CV-MLP achieves in average ~2% more accuracy than RV-MLP
- CVNN took more epochs to converge
- Non-overlapping results
  - CV-MLP minimum value was higher than RV-MLP maximum value