Neural Transfer Function Library Deriving From Parametric Probability Signatures

Abdourrahmane Mahamane Atto

Neural Transfer Function Library Deriving From Parametric Probability Signatures

Abdourrahmane Mahamane ATTO*

***Abstract*—This letter highlights new artificial neuron activation models designed with respect to some major parametric probability distribution families, in light of the paper [1].**

In what follows, sgn denotes the sign function, $\mathbb{1}_{\mathcal{E}}$ is the indicator of set $\mathcal{E}$ and "$\triangleq$" represents the *denotation* symbol. The *Neural Activation* (NA) forms (or *transfer functions*) provided in [1] include the *Generalized Normal* NA (GNNA):

$$\mathrm{GNNA}_{\mu,\sigma,\beta}(x) = \frac{x\mathbb{1}_{\{x\geqslant 0\}}}{1+e^{-\operatorname{sgn}(x-\mu)\left(\frac{|x-\mu|}{\sigma}\right)^{\beta}}} \tag{1}$$

having a *location* or *shift* parameter $\mu \geqslant 0$, a *scale* parameter $\sigma > 0$ and a *shape* parameter $\beta > 0$. GNNA family includes:

- the *Laplace* Neural Activation (LNA) when $\beta = 1$:

$$\mathrm{LNA}_{\mu,\sigma}(x) = \frac{x\mathbb{1}_{\{x\geqslant 0\}}}{1+e^{-\left(\frac{x-\mu}{\sigma}\right)}} \tag{2}$$

- the *Normal* Neural Activation (NNA) when $\beta = 2$:

$$\mathrm{NNA}_{\mu,\sigma}(x) = \frac{x\mathbb{1}_{\{x\geqslant 0\}}}{1+e^{-\operatorname{sgn}(x-\mu)\left(\frac{x-\mu}{\sigma}\right)^{2}}} \tag{3}$$

Note also that when μ is chosen to be 0, then (1), (2) and (3) reduce to neural transfer functions that can be called respectively Generalized Half-Normal (GHNNA), Half-Laplace (HLNA) and Half-Normal (HNNA).

From the underlying formulation principle in (1), we derive a *Generalized Gamma* NA (GGNA) with *scale* parameter $\sigma > 0$ and two *shape* parameters $\beta_1, \beta_2 > 0$ when specifying:

$$\mathrm{GGNA}_{\sigma,\beta_1,\beta_2}(x) = \frac{x\mathbb{1}_{\{x\geqslant 0\}}}{1+\left(\frac{x}{\sigma}\right)^{\beta_2-1} e^{-\left(\frac{x}{\sigma}\right)^{\beta_1}}} \tag{4}$$

This GGNA family includes:

- the *Weibull* NA (WNA) when $\beta_1 = \beta_2 \triangleq \beta$:

$$\mathrm{WNA}_{\sigma,\beta}(x) = \frac{x\mathbb{1}_{\{x\geqslant 0\}}}{1+\left(\frac{x}{\sigma}\right)^{\beta-1} e^{-\left(\frac{x}{\sigma}\right)^{\beta}}} \tag{5}$$

- the *Gamma* NA (GNA) when $\beta_1 = 1$ and $\beta_2 \triangleq \beta$:

$$\mathrm{GNA}_{\sigma,\beta_1,\beta_2}(x) = \frac{x\mathbb{1}_{\{x\geqslant 0\}}}{1+\left(\frac{x}{\sigma}\right)^{\beta-1} e^{-\left(\frac{x}{\sigma}\right)}} \tag{6}$$

- the *Exponential* NA (ENA) when $\beta_1 = \beta_2 = 1$:

$$\mathrm{ENA}_{\sigma}(x) = \frac{x\mathbb{1}_{\{x\geqslant 0\}}}{1+e^{-\frac{x}{\sigma}}} \tag{7}$$

- the *Rayleigh* NA (RNA) when $\beta_1 = \beta_2 = 2$:

$$\mathrm{RNA}_{\sigma}(x) = \frac{x\mathbb{1}_{\{x\geqslant 0\}}}{1+\left(\frac{x}{\sigma}\right) e^{-\left(\frac{x}{\sigma}\right)^{2}}} \tag{8}$$

Another formulation concerns the definition of a *Multivariate Normal* NA (MNNA) associated with (3) from the specification:

$$\mathrm{MNNA}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(x_k) = \frac{x_k\mathbb{1}_{\{x_k\geqslant 0\}}}{1+e^{-(\operatorname{sgn}(x_k-\mu_k))(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}(\mathbf{x}-\boldsymbol{\mu})'}} \tag{9}$$

where the input is a row vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_N)$ and the learnable activation variables are:

- the location parameters $\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_N)$, a vector of non-negative value, and
- a symmetric positive definite *scale* matrix of parameters $\boldsymbol{\Sigma} = (\sigma_{k,\ell})_{1\leqslant k,\ell,\leqslant N}$.

Extension of (9) to *Multivariate Elliptical* NA (MENA) is then a straightforward replacement of the exponential term used in (9) to derive a more general form:

$$\mathrm{MENA}_{\boldsymbol{\mu},\boldsymbol{\Sigma}}(x_k) = \frac{x_k\mathbb{1}_{\{x_k\geqslant 0\}}}{1+g\left(\epsilon_k\,(\mathbf{x}-\boldsymbol{\mu})\boldsymbol{\Sigma}(\mathbf{x}-\boldsymbol{\mu})'\right)} \tag{10}$$

where $\epsilon_k = \operatorname{sgn}(x_k - \mu_k)$. We can then switch from the general formula (10) to different multivariate neural transfer functions thanks to convenient selections of the so-called elliptical *generator* g given in (10). For instance, $g(z) = e^{-z}$ leads to MNNA of (9) whereas the *Multivariate Student* NA (MSNA) follows by using $g_\nu(z) = (1 + z/\nu)^{-(N+\nu)/2}$, a *Multivariate Logistic* NA (MLNA) can be derived from $g(z) = e^{-z}/\left(1+e^{-z}\right)^2$ and a *Multivariate Kotz* NA (MKNA) can be specified thanks to $g_{\nu_1,\nu_2,\nu 3}(z) = z^{\nu_1-1}e^{-\nu_2 z^{\nu_3}}$.

One can note that functions pertaining to GNNA, GGNA and MENA families perform a shrinkage on their input neural information (which corresponds in practice to the aggregation of post-synaptic signals). But similarly to [1], we can adjoin to these functions, some complementary stretchage neural information forms defined as $2x\mathbb{1}_{\{x\geqslant 0\}} - f_\Theta(x)$ where f_Θ is one among (1), (2), (3), (4), (5), (6), (7), (8) or (9) neural activations.

We end this comment paper by noting that in all the aforementioned neural activation models, the normalization constants of the underlying probability model signatures are considered as superfluous because: (i) certain constants can be integrated in the learnable parameters and (ii) the "unit sum" normalization constants often introduce special functions and the derivatives of these functions with respect to learnable parameters can lead to numerical instabilities.

* Université Savoie Mont Blanc - LISTIC

References

[1] A. M. Atto, S. Galichet, D. Pastor, and N. Méger, "On joint parameterizations of linear and nonlinear functionals in neural networks," *Neural Networks*, vol. 160, pp. 12–21, 2023. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0893608022005111 1