Diseño de Atributos
desde
Serie de Tiempo
+
Red Neurona Superficial (SNN).

Prof. NIBALDO RODRÍGUEZ A.

Descomposición de Series de Tiempo (ST)

- Sea ***X*** una serie de tiempo discreta de longitud ***N***, dado por:

$$\boxed{\mathrm{X} = \{\mathrm{x}_\mathrm{n}\}_{n=1}^{N}}$$

$$\boxed{\mathrm{X} = \{x_1, x_2, \cdots x_L, \ldots, x_K, \ldots, x_N\}}$$

- Descomposición en L-componentes intrínsecos desde una ST:

$$\boxed{\mathrm{X} = \sum_{i=1}^{L} C_i}$$

Algoritmo de Extracción de Componentes desde ST:

- Paso #1: Construir Matriz de Hankel.
- Paso #2: SVD de Matriz Hankel.
- Paso #3: Calcular Componentes.

Matriz de Hankel:

$$X = \{x_1, x_2, \cdots x_L, \ldots, x_K, \ldots, x_N\}$$

$$H = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & \ldots & x_K \\ x_2 & x_3 & x_4 & x_5 & \ldots & x_{K+1} \\ x_3 & x_4 & x_5 & x_6 & \ldots & x_{K+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \ldots & \ldots & x_N \end{pmatrix}$$

$$K = N - L + 1$$

MATRIZ HANKEL

$$\boxed{\mathrm{X} = \{x_1, x_2, \cdots, x_N\}}$$

$$H = \begin{pmatrix} x_1 & x_2 & x_3 & x_4 & \dots & x_K \\ x_2 & x_3 & x_4 & x_5 & \dots & x_{K+1} \\ x_3 & x_4 & x_5 & x_6 & \dots & x_{K+2} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_L & x_{L+1} & x_{L+2} & \dots & \dots & x_N \end{pmatrix}$$

Ejemplo: MATRIZ HANKEL

$$\mathrm{X} = \{\mathrm{x}_{\mathrm{n}}\}_{n=1}^{N},\ N = 12$$

$$X = \{1,2,3,\ldots,10,11,12\}$$

$$L = 5$$

$$\mathrm{H} = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 & 9 \\ 3 & 4 & 5 & 6 & 7 & 8 & 9 & 10 \\ 4 & 5 & 6 & 7 & 8 & 9 & 10 & 11 \\ 5 & 6 & 7 & 8 & 9 & 10 & 11 & 12 \end{pmatrix}$$

Extracción de Componentes usando H-SVD:

- Paso #1: Calcular matriz Hankel: L-filas por K-columna

$$\boxed{H = matriz(L, K)}$$

- Paso #2: Descomposición de valores singular de Hankel :

$$\boxed{H = U \times S \times V^T, \text{ donde}}$$

$$\boxed{H = H_1 + H_2 + \ldots, H_L}$$

Extracción de Componentes usando H-SVD

- Paso #3: Calcular componentes de Hankel:

$$H_i = s(i) \times U(:,i) \times V(:,i)^T,$$
$$i = 1,2,\ldots,L,$$
$$s = diag(S)$$

Extracción de Componentes usando H-SVD

- Paso #3: Calcular componentes de Hankel:

$$H_i = \begin{pmatrix} x_1^i & x_2^i & x_3^i & x_4^i & \dots & x_K^i \\ x_2^i & x_3^i & x_4^i & x_5^i & \dots & x_{K+1}^i \\ x_3^i & x_4^i & x_5^i & x_6^i & \dots & x_{K+2}^i \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_L^i & x_{L+1}^i & x_{L+2}^i & \dots & \dots & x_N^i \end{pmatrix}$$

$$\mathrm{C}_{\mathrm{i}} = \left[x_1^i \quad x_2^i \quad \dots \quad x_K^i \quad x_{k+1}^i \quad \dots \quad x_N^i \right],$$
$$i = 1, \dots, L$$

Componente *i–th* de Hankel

Ejemplo: H-SVD

$$X = [1.2371 \quad 0.1797 \quad 0.0042 \quad 1.5944 \quad 0.7729 \quad 1.0414 \quad 0.0035]$$

$$L - row = 3$$

$$\boxed{H = \begin{bmatrix} 1.2371 & 0.1797 & 0.0042 & 1.5944 & 0.7729 \\ 0.1797 & 0.0042 & 1.5944 & 0.7729 & 1.0414 \\ 0.0042 & 1.5944 & 0.7729 & 1.0414 & 0.0035 \end{bmatrix}}$$

Ejemplo: H-SVD

$$U = \begin{bmatrix} -0.6100 & 0.6912 & 0.3875 \\ -0.5762 & -0.0512 & -0.8157 \\ -0.5439 & -0.7209 & 0.4295 \end{bmatrix}$$

$$sv = \begin{bmatrix} 2.9308 & 1.5786 & 1.4493 \end{bmatrix}$$

$$V = \begin{bmatrix} -0.2936 & 0.5339 & 0.2309 \\ -0.3341 & -0.6495 & 0.5182 \\ -0.4578 & -0.4028 & -0.6671 \\ -0.6771 & 0.1975 & 0.2999 \\ -0.3663 & 0.3030 & -0.3784 \end{bmatrix}$$

Ejemplo: H-SVD

$$H_1 = \begin{bmatrix} 0.5249 & 0.5974 & 0.8184 & 1.2105 & 0.6548 \\ 0.4958 & 0.5643 & 0.7731 & 1.1434 & 0.6185 \\ 0.4680 & 0.5327 & 0.7298 & 1.0794 & 0.5839 \end{bmatrix}$$

$$C_1 = \begin{bmatrix} 0.5249 & 0.5974 & 0.8184 & 1.2105 & 0.6548 & 0.6185 & 0.5839 \end{bmatrix}$$

Ejemplo: H-SVD

$$H_2 = \begin{bmatrix} 0.5825 & -0.7087 & -0.4395 & 0.2155 & 0.3307 \\ -0.0432 & 0.0525 & 0.0326 & -0.0160 & -0.0245 \\ -0.6075 & 0.7391 & 0.4584 & -0.2247 & -0.3448 \end{bmatrix}$$

$$C_2 = \begin{bmatrix} 0.5825 & -0.7087 & -0.4395 & 0.2155 & 0.3307 & -0.0245 & -0.3448 \end{bmatrix}$$

Ejemplo: H-SVD

$$H_3 = \begin{bmatrix} 0.1297 & 0.2910 & -0.3747 & 0.1684 & -0.2125 \\ -0.2730 & -0.6126 & 0.7887 & -0.3546 & 0.4473 \\ 0.1437 & 0.3226 & -0.4153 & 0.1867 & -0.2356 \end{bmatrix}$$

$$C_3 = \begin{bmatrix} 0.1297 & 0.2910 & -0.3747 & 0.1684 & -0.2125 & 0.4473 & -0.2356 \end{bmatrix}$$

Ejemplo: H-SVD

$$C_1 = [0.5249 \quad 0.5974 \quad 0.8184 \quad 1.2105 \quad 0.6548 \quad 0.6185 \quad 0.5839]$$

$$C_2 = [0.5825 \quad -0.7087 \quad -0.4395 \quad 0.2155 \quad 0.3307 \quad -0.0245 \quad -0.3448]$$

$$C_3 = [0.1297 \quad 0.2910 \quad -0.3747 \quad 0.1684 \quad -0.2125 \quad 0.4473 \quad -0.2356]$$

$$X = C_1 + C_2 + C_3$$

Series de tiempo:

Descomposición Diádica:

Prof. NIBALDO RODRÍGUEZ A.

Descomposición Diádica

- Considere una serie de tiempo discreta : $\boxed{\mathrm{X} = \{\mathrm{x}_{\mathrm{n}}\}_{n=1}^{N}}$

Matriz de Hankel:

$$H = \begin{pmatrix} x_1 & x_2 & x_3 & \cdots & x_{N-1} \\ x_2 & x_3 & x_4 & \cdots & x_N \end{pmatrix}$$

Descomposición diádica:

$$[U \;\; \mathrm{S} \;\; \mathrm{V}] = SVD(H)$$
$$s = diag(S)$$

Hankelización:

$$H_1 = s(1) \times U(:,1) \times V(:,1)^T$$
$$H_2 = s(2) \times U(:,2) \times V(:,2)^T$$

Descomposición Diádica

- **Componentes Diádicos:**

$$\boxed{H_1 = \begin{pmatrix} x_1^1 & x_2^1 & x_3^1 & \dots & x_{N-1}^1 \\ x_2^1 & x_3^1 & x_4^1 & \dots & x_N^1 \end{pmatrix}}$$

$$\boxed{C_1 = \begin{bmatrix} x_1^1 & \overline{x}_2^1 & \overline{x}_3^1 & \dots & \overline{x}_{N-1}^1 & x_N^1 \end{bmatrix}}$$

Descomposición Diádica

- **Componentes Diádicos:**

$$\boxed{H_2 = \begin{pmatrix} x_1^2 & x_2^2 & x_3^2 & \dots & x_{N-1}^2 \\ x_2^2 & x_3^2 & x_4^2 & \dots & x_N^2 \end{pmatrix}}$$

$$\boxed{C_2 = \begin{bmatrix} x_1^2 & \overline{x}_2^2 & \overline{x}_3^2 & \dots & \overline{x}_{N-1}^2 & x_N^2 \end{bmatrix}}$$

$$\boxed{\mathrm{X} = C_1 + C_2}$$

Ejemplo: Descomposición Diádica

- Sea x= {1,2,…7}

$$H = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ 2 & 3 & 4 & 5 & 6 & 7 \end{bmatrix}$$

- Descomposición Diádica:

$$[U \;\; \mathrm{S} \;\; \mathrm{V}] = SVD(H)$$

$$V = \begin{bmatrix} -0.1442 & 0.7092 \\ -0.2370 & 0.4890 \\ -0.3298 & 0.2687 \\ -0.4226 & 0.0484 \\ -0.5154 & -0.1719 \\ -0.6083 & -0.3922 \end{bmatrix}$$

$$U = \begin{bmatrix} -0.6287 & -0.7777 \\ -0.7777 & 0.6287 \end{bmatrix}$$

$$s = diag(S) = [15.1507 \;\; 0.6763]$$

Ejemplo: Descomposición Diádica

- Hankelización:

$$H_1 = s(1) \times U(:,1) \times V(:,1)^T$$

$$H_1 = \begin{bmatrix} 1.3730 & 2.2572 & 3.1413 & 4.0254 & 4.9096 & 5.7937 \\ 1.6984 & 2.7921 & 3.8858 & 4.9794 & 6.0731 & 7.1668 \end{bmatrix}$$

- Componentes Diádicos::

$$C_1 = \begin{bmatrix} 1.3730 & 1.9778 & 2.9667 & 3.9556 & 4.9445 & 5.9334 & 7.1668 \end{bmatrix}$$

Ejemplo: Descomposición Diádica

- Hankelización:

$$H_2 = s(2) \times U(:,2) \times V(:,2)^T$$

$$H_2 = \begin{bmatrix} -0.3730 & -0.2572 & -0.1413 & -0.0254 & 0.0904 & 0.2063 \\ 0.3016 & 0.2079 & 0.1142 & 0.0206 & -0.0731 & -0.1668 \end{bmatrix}$$

- Componentes Diádicos::

$$C_2 = [-0.3730 \quad 0.0222 \quad 0.0333 \quad 0.0444 \quad 0.0555 \quad 0.0666 \quad -0.1668]$$

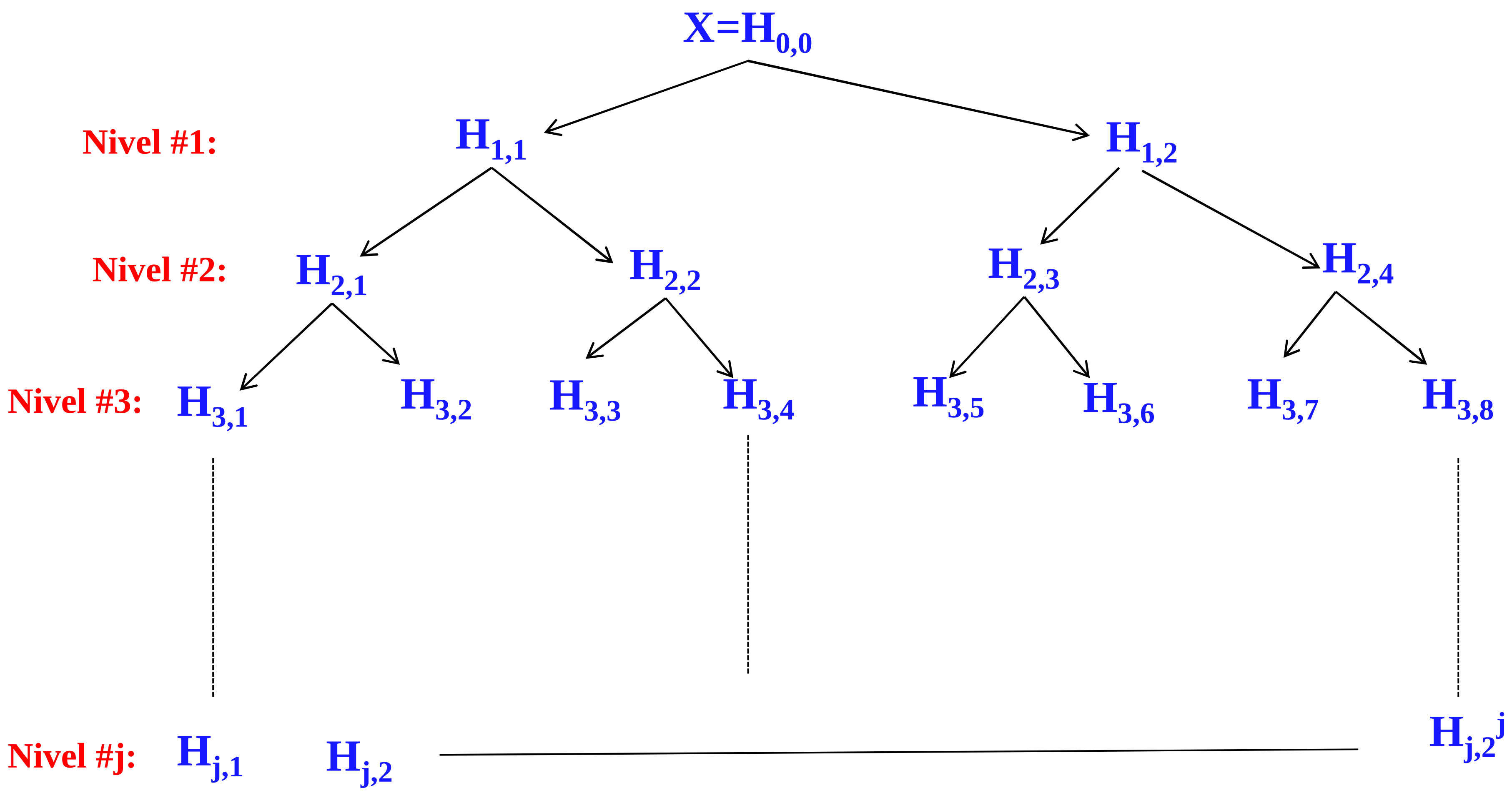
$$\boxed{x = C_1 + C_2}$$

Series de tiempo:

Descomposición Multi-escala Diádica usando Hankel

Prof. NIBALDO RODRÍGUEZ A.

Descomposición Multi-escala



Máx. Niveles: $j = \log_2(N)$

Ejemplo: Descomposición Multi-scale

- Sea x= {1,2,…8}

$$H = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix}$$

- Nivel #1:

$$C_{1,1} = [1.7448 \;\; 1.1712 \;\; 1.0361 \;\; 1.0806 \;\; 0.7769 \;\; 0.7084 \;\; 0.7606 \;\; 0.6123]$$

$$C_{1,2} = [0.233 \;\; -0.225 \;\; -0.076 \;\; 0.313 \;\; -0.123 \;\; -0.208 \;\; 0.5145 \;\; -0.594]$$

Ejemplo: Descomposición Multi-scale

- Sea x= {1,2,…8}

$$H = \begin{bmatrix} 1 & 2 & 3 & 4 & 5 & 6 & 7 \\ 2 & 3 & 4 & 5 & 6 & 7 & 8 \end{bmatrix}$$

- Nivel #2:

$$C = \begin{bmatrix} 1.61 & 1.26 & 1.07 & 0.98 & 0.82 & 0.73 & 0.70 & 0.6200 \\ 0.12 & -0.09 & -0.03 & 0.09 & -0.05 & -0.02 & 0.05 & -0.0077 \\ 0.19 & -0.15 & -0.05 & 0.20 & -0.08 & -0.19 & 0.44 & -0.6308 \\ 0.04 & -0.07 & -0.01 & 0.10 & -0.03 & -0.01 & 0.06 & 0.0360 \end{bmatrix}$$

Ejemplo: Descomposición Multi-scale

- Nivel #3:

$$C=\begin{bmatrix} 1.56 & 1.296 & 1.091 & 0.960 & 0.837 & 0.744 & 0.686 & 0.608 \\ 0.05 & -0.029 & -0.021 & 0.024 & -0.009 & -0.012 & 0.0187 & 0.0123 \\ 0.14 & -0.07 & -0.017 & 0.067 & -0.042 & -0.013 & 0.0351 & -0.028 \\ -0.007 & -0.027 & -0.017 & 0.029 & -0.009 & -0.010 & 0.021 & 0.020 \\ 0.125 & -0.105 & -0.042 & 0.136 & -0.045 & -0.181 & 0.414 & -0.649 \\ 0.06 & -0.04 & -0.017 & 0.069 & -0.042 & -0.0161 & 0.0355 & 0.0184 \\ 0.06 & -0.04 & -0.017 & 0.068 & -0.043 & -0.0126 & 0.0264 & -0.0155 \\ -0.01 & -0.03 & -0.0002 & 0.041 & 0.006 & 0.0017 & 0.0390 & 0.0515 \end{bmatrix}$$

Cuantificación de la Información

Entropía de Shannon

- Permite medir el contenido de información de un conjunto de datos.
- Permite medir la Incertidumbre de un conjunto de datos.

$$\text{Sea} \quad X = \{x_1, x_2, \ldots, x_N\}$$

$$H(X) = -\sum_{i=1}^{N} P(x_i) \log_2 \left(P(x_i)\right)$$

$$H(X) = -\frac{1}{\log_2 (N)} \sum_{i=1}^{N} P(x_i) \log_2 \left(P(x_i)\right)$$

Entropía Normalizada

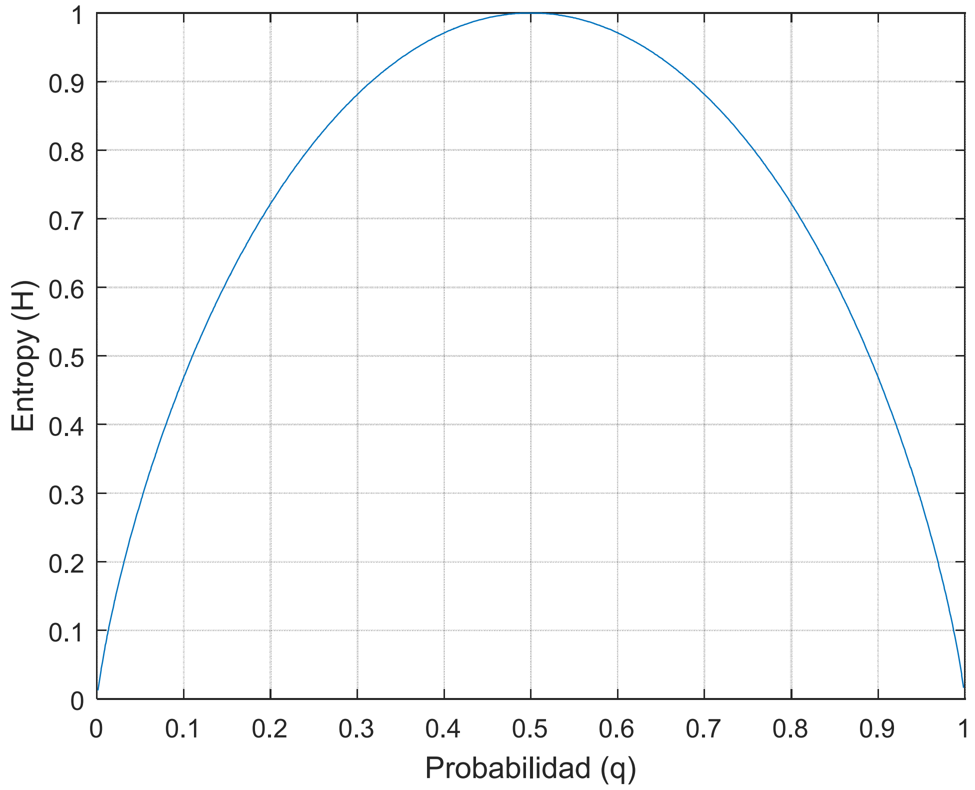
Ejemplo: Entropía de Shannon

- Sea un conjunto de eventos **$X = \{ x_1, x_2 \}$, donde:**

$$\boxed{\begin{array}{ll} P(x_1) = a, & a \in [0,1] \\ P(x_2) = 1 - a & \end{array}}$$

$$\boxed{H(X) = -\left\{ P(x_1)\log_2\left(P(x_1)\right) + P(x_2)\log_2\left(P(x_2)\right) \right\}}$$

Ejemplo: Entropía de Shannon



- A mayor Entropía menos Información

Entropía Espectral de Shannon:

Transformada de Fourier

Prof. NIBALDO RODRÍGUEZ A.

Transformada de Fourier continua

- Transformada de Fourier Continua (CFT):

$$F(f) = \int_{-\infty}^{\infty} x(t) e^{-j2\pi ft} dt$$

$$j = \sqrt{-1}$$

- Transformada de Fourier continua Inversa (ICFT):

$$x(t) = \int_{-\infty}^{\infty} F(f) e^{j2\pi ft} df$$

- **t**: denota el tiempo de la señal, ***f***: representa la frecuencia de Fourier

Descomposición de Fourier

- Transformada de Fourier Discreta (DFT):

$$F(k) = \frac{1}{N}\sum_{n=0}^{N-1} x(n)\exp\left(-j\frac{2\pi}{N}kn\right)$$

$$k = 0,1,\ldots,N-1, \quad j = \sqrt{-1}$$

- Transformada de Fourier Inversa Discreta (IDFT):

$$x(n) = N \times \sum_{k=0}^{N-1} F(k)\exp\left(j\frac{2\pi}{N}kn\right)$$

$$n = 0,1,\ldots,N-1$$

- **N**: número de muestras de la serie de tiempo, *k*: k-th línea de frecuencia de Fourier.

Espectro de Fourier

$$F(k) = A(k) \times \exp\left(j\Phi(k)\right)$$

$$F(k) = a_k + jb_k$$

$$A(k) = \sqrt{a_k^2 + b_k^2}$$

$$k = 0,1,2,\ldots,N-1$$

Ejemplo: FT de Función Coseno

$$f(t) = \cos(2\pi \times f_0 \times t)$$

$$f(t) = \frac{1}{2}\left[e^{j2\pi f_0 t} + e^{-j2\pi f_0 t}\right]$$

$$F(f) = \int_{-\infty}^{\infty} f(t) e^{-j2\pi ft} dt$$

$$F(f) = \frac{1}{2}\int_{-\infty}^{\infty} e^{-j2\pi(f-f_0)t} dt + \frac{1}{2}\int_{-\infty}^{\infty} e^{-j2\pi(f+f_0)t} dt$$

$$F(f) = \frac{1}{2}\left[\delta(f - f_0) + \delta(f + f_0)\right]$$

$$\delta(x) = \begin{cases} 1 & x = 0 \\ 0 & x \neq 0 \end{cases}$$

Ejemplo: FT de Función Seno

$$f(t) = \sin\left(2\pi \times f_0 \times t\right)$$

$$f(t) = \frac{1}{2j}\left[e^{j2\pi f_0 t} - e^{-j2\pi f_0 t}\right]$$

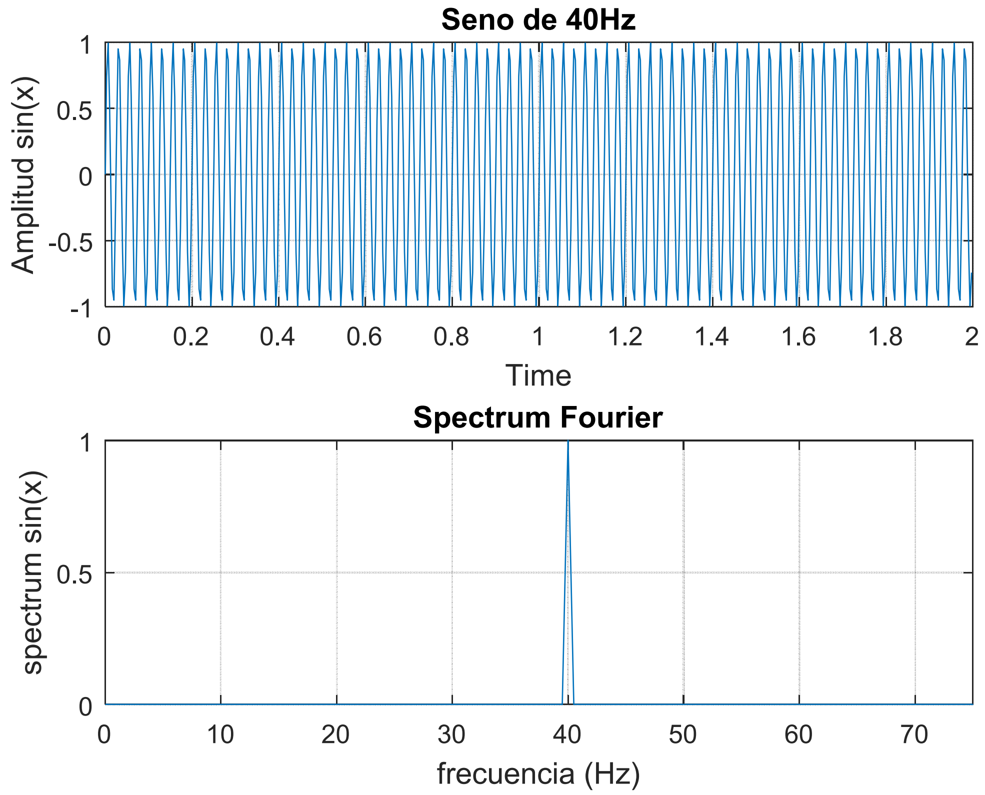
$$F(f) = \int_{-\infty}^{\infty} f(t) e^{-j2\pi ft} dt$$

$$F(f) = \frac{1}{2j}\int_{-\infty}^{\infty} e^{-j2\pi(f-f_0)t} dt - \frac{1}{2j}\int_{-\infty}^{\infty} e^{-j2\pi(f+f_0)t} dt$$

$$F(f) = -\frac{j}{2}\delta(f - f_0) + \frac{j}{2}\delta(f + f_0)$$

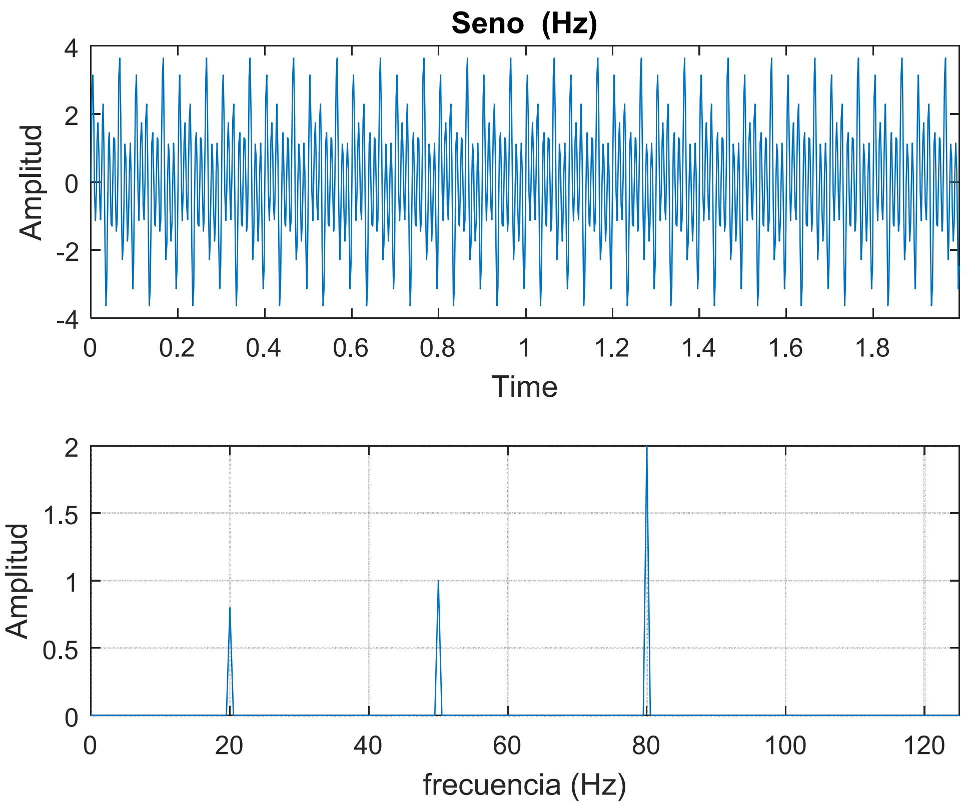
Ejemplo: DFT de Seno

$$f(t) = \sin(2\pi 40t), \quad Fs = 300\left(\frac{\text{muestras}}{\text{sec}}\right)$$



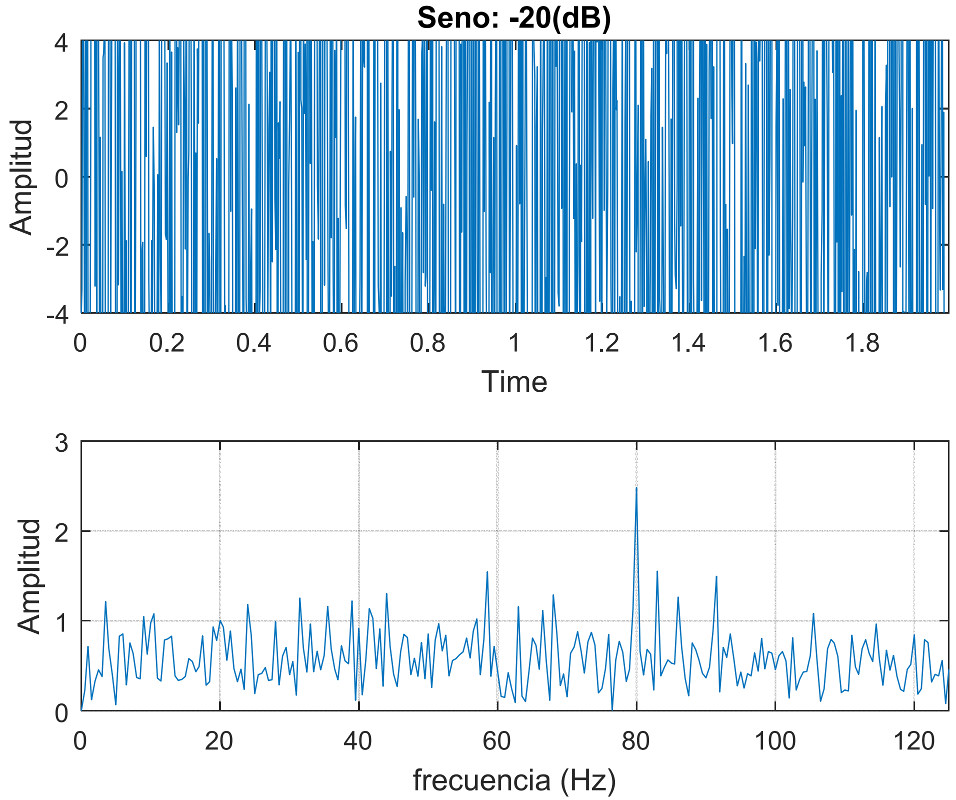
Ejemplo: DFT de Seno

$$f(t) = 0.8\sin(2\pi 20t) + \sin(2\pi 50t) + \sin(2\pi 80t), \quad Fs = 500\left(\frac{\text{muestras}}{\text{sec}}\right)$$



Ejemplo: DFT de Seno

$$f(t) = 0.8\sin(2\pi 20t) + \sin(2\pi 50t) + \sin(2\pi 80t) + n(t), \quad Fs = 500\left(\frac{\text{muestras}}{\text{sec}}\right)$$



Entropía Espectral de Shannon

- Sea **X una serie de tiempo ={ x_1, x_2,…,x_N}.**
- **La Entropía Espectral es calculada como:**

$$H(X) = -\frac{1}{\log_2(N)}\sum_{k=1}^{N} P(x_k)\log_2\left(P(x_k)\right)$$

$$P(x_k) = \frac{A_k^2}{\sum_{k=1}^{N} A_k^2}$$

$$A_k = \left|\sum_{n=1}^{N} x_n \exp\left(-j2\pi\times\frac{k\times n}{N}\right)\right|$$

Red Neuronal Superficial (SNN)

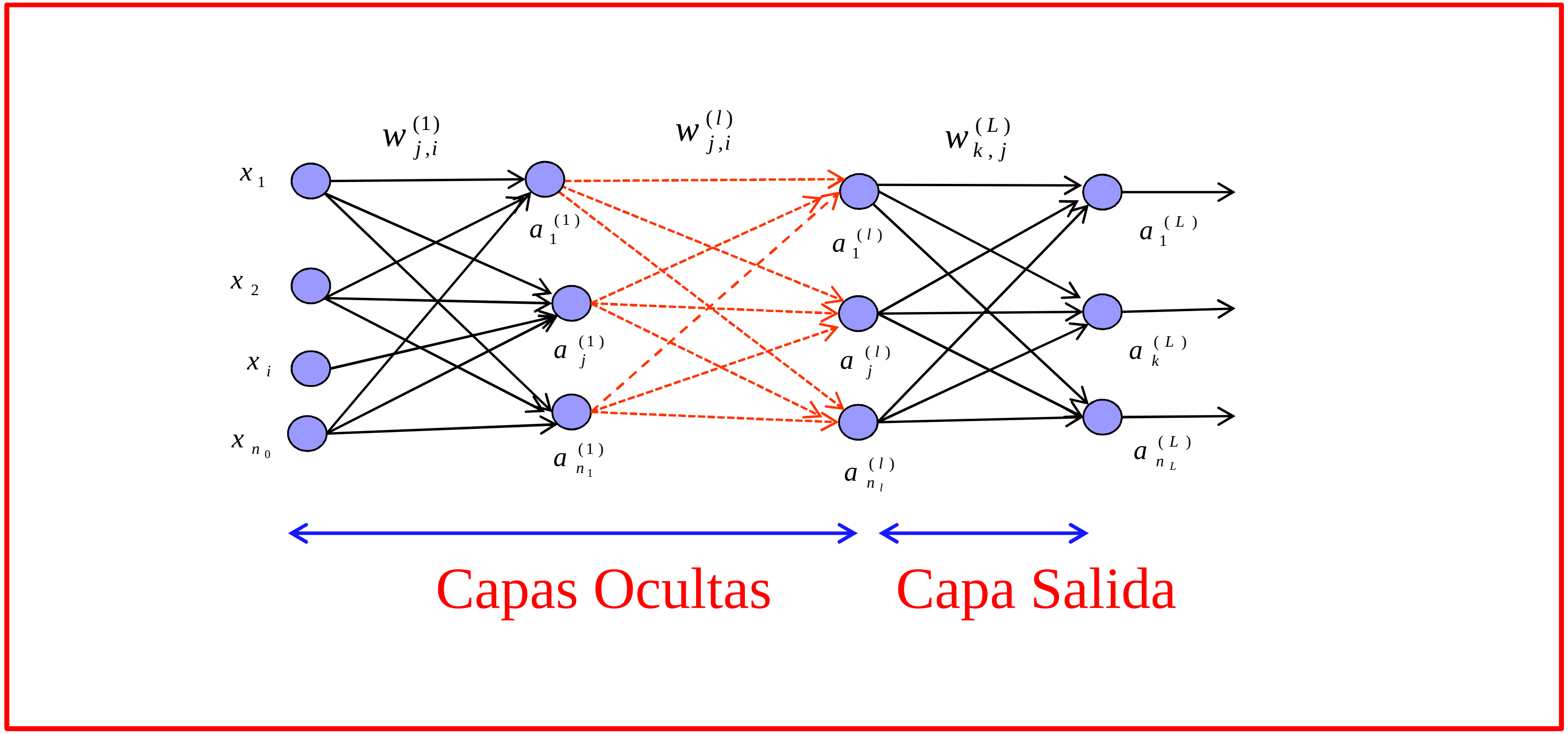
Algoritmo de Aprendizaje:

miniBatch RMSprop

+

miniBatch Adam

SNN



$w_{j,i}^{(1)}$
$w_{j,i}^{(l)}$
$w_{k,j}^{(L)}$
x_1
x_2
x_i
x_{n_0}
$a_1^{(1)}$
$a_j^{(1)}$
$a_{n_1}^{(1)}$
$a_1^{(l)}$
$a_j^{(l)}$
$a_{n_l}^{(l)}$
$a_1^{(L)}$
$a_k^{(L)}$
$a_{n_L}^{(L)}$
Capas Ocultas
Capa Salida

Algoritmo de Aprendizaje: miniBatch RMSprop

Training de SNN: miniBatch RMSprop

- Considerar una base de datos de *N*-muestras dada como:

$$\boxed{\{X_i, Y_i\}_{i=1}^{N}, \quad X_i \in \Re^d, \; Y_i \in \Re^m}$$

- X_i : representa la data de training
- Y_i: representa la data deseada

miniBatch RMSprop: Etapas del Algoritmo

- **Paso 1**: Re-ordenar aleatoriamente la localización de cada muestra de la base de datos de training.

- **Paso 2**: Dividir las ***N***-muestras de datos en Batch de ***M***-muestras.

$$T = \frac{N}{M}$$

T : Número de Batch

- **Paso 3**: Entrenar la SNN usando un Número Máximo de Épocas.
  - Cada Épocas ajusta los pesos de la SNN ***T***-veces vía el RMSProp.

Algoritmo Aprendizaje: miniBatch RMSprop

Costo: MSE $$\boxed{E = \frac{1}{2M}\sum_{n=1}^{M} C_n = \frac{1}{2M}\sum_{n=1}^{M}\sum_{k=1}^{n_L}\left(a_{k,n}^{(L)} - d_{k,n}\right)^2}$$

Activación Capa Salida: $$\boxed{a^{(L)} = f\left(a^{(L-1)}\right)}$$

Activación Capas Ocultas: $$\boxed{a^{(l)} = g\left(a^{(l-1)}\right),\ l = 1,2,\ldots L-1}$$

- **M: número de muestras , n_L: nodos de salidas, $d_{k,n}$: valor deseado**
- **f(.): función de activación Capa de Salida**
- **g(.): función de activación Capas Ocultas**

Ajuste Pesos de Salida: miniBatch RMSprop

Pesos de Salida:

$$W^{(L)}(t) = W^{(L)}(t-1) - \mu \times gRMS$$
$$\mu \in (0,1),\ t = 1, \dots, T$$

Notación Matricial:
Actualización Pesos Salida

$$gRMS = \frac{1}{\sqrt{V^{(L)}(t) + \varepsilon}} \otimes gW^{(L)}$$
$$V^{(L)}(t) = \beta \times V^{(L)}(t-1) + (1-\beta) \times \left(gW^{(L)}\right)^2$$

Notación Matricial:
Gradiente

$$gW^{(L)} = \delta^{(L)} \times \left(a^{(L-1)}\right)^T$$
$$\delta^{(L)} = e^{(L)} \otimes f'\left(z^{(L)}\right)$$

Valores Iniciales:

$$\beta = 0.9, \varepsilon \in \left(10^{-6}, 10^{-10}\right)$$
$$V^{(L)} = W^{(L)}.shape$$
$$V^{(L)}(0) = 0$$

Ajuste Pesos Ocultos: SNN+RMSprop

Pesos de Ocultos:

$$W^{(l)}(t) = W^{(l)}(t-1) - \mu \times gRMS$$

Notación Matricial:
Actualización Pesos Ocultos

$$gRMS = \frac{1}{\sqrt{V^{(l)}(t) + \varepsilon}} \otimes gW^{(l)}$$

$$V^{(l)}(t) = \beta \times V^{(l)}(t-1) + (1-\beta) \times \left(gW^{(l)}\right)^2$$

Notación Matricial:
Gradiente

$$gW^{(l)} = \left\{\left(w^{(l+1)}\right)^T \times \delta^{(l+1)}\right\} \otimes g'\left(z^{(l)}\right) \times \left(a^{(l-1)}\right)^T$$
$$l = L-1, L-2, \ldots, 3, 2, 1$$

Valores Iniciales:

$$\beta = 0.9, \varepsilon \in \left(10^{-6}, 10^{-10}\right)$$
$$V^{(l)} = W^{(l)}.shape, \; l = 1, 2, \ldots L-1$$
$$V^{(l)}(0) = 0$$

Red Neuronal Superficial (SNN):

Algoritmo de Aprendizaje miniBatch Adam

Algoritmo Aprendizaje: miniBatch Adam

Costo:
Entropía cruzada

$$E = -\frac{1}{M}\sum_{n=1}^{M}\sum_{k=1}^{n_L} d_{k,n}\log\left(a_{k,n}^{(L)}\right)$$

N = Número muestras

n_L = Nodos de salidas

$d_{k,n}$ = Valor deseado de la clase k.

Activación Capa Salida:

$$a_n^{(L)} = \frac{z^{(L)}(:,n)}{\sum_{k=1}^{n_L} z_{k,n}^{(L)}}, \quad n = 1,\dots,M$$

$$z^{(L)} = \exp\left(w^{(L)} \times a^{(L-1)}\right), \quad \mathrm{z}^{(\mathrm{L})} \in \Re^{(n_L \times M)}, w^{(L)} \in \Re^{(n_L \times n_{L-1})}$$

Activación Capas Ocultas:

$$a^{/l)} = g(a^{(l-1)}),$$
$$l = L-1, L-2, \dots, 1$$

Algorithm Aprendizaje: miniBatch Adam

Actualización Pesos de Salida

$$\boxed{\begin{aligned} & w^{(L)}(t) = w^{(L)}(t-1) - \mu \times gAdam \\ & t = 1,2,\ldots, MaxIter \\ & u \in (0,1) \end{aligned}}$$

Algorithm Aprendizaje: miniBatch Adam

Actualización Pesos de Salida

$$gAdam = \frac{\sqrt{1-\beta_2^t}}{1-\beta_1^t} \times \frac{V^{(L)}(t)}{\sqrt{S^{(L)}(t)+\varepsilon}}, \qquad \varepsilon = \left(10^{-6}, 10^{-8}\right)$$

$$V^{(L)}(t) = \beta_1 \times V^{(L)}(t-1) + (1-\beta_1)\frac{\partial E}{\partial w^{(L)}(t-1)}$$

$$S^{(L)}(t) = \beta_2 \times S^{(L)}(t-1) + (1-\beta_2)\left(\frac{\partial E}{\partial w^{(L)}(t-1)}\right)^2$$

$$w^{(L)}(0) = \text{random}$$

$$V^{(L)}(0),\ S^{(L)}(0) = W^{(L)}.shape$$

$$V^{(L)}(0),\ S^{(L)}(0) = 0$$

$$\beta_1 = 0.9 \qquad \beta_2 = 0.999$$

Algoritmo Aprendizaje: miniBatch Adam

Actualización Pesos Ocultos

$$\boxed{\begin{array}{l} w^{(l)}(t) = w^{(l)}(t-1) - \mu \times gAdam \\ l = L-1, L-2, \ldots\ 2,1 \\ t = 1,2,\ldots\ , MaxIter \\ \mu \in (0,1) \end{array}}$$

Algorithm Aprendizaje: miniBatch Adam

Actualización Pesos Ocultos

$$gAdam = \frac{\sqrt{1-\beta_2^t}}{1-\beta_1^t} \times \frac{V^{(l)}(t)}{\sqrt{S^{(l)}(t)+\varepsilon}}$$

$$V^{(l)}(t) = \beta_1 V^{(l)}(t-1) + (1-\beta_1)\frac{\partial E}{\partial w^{(l)}(t-1)}$$

$$S^{(l)}(t) = \beta_2 S^{(l)}(t-1) + (1-\beta_2)\left(\frac{\partial E}{\partial w^{(l)}(t-1)}\right)^2$$

$$w^{(l)}(0) = \text{random} \quad , \; l = 1,2,\ldots, L-1$$

$$V^{(l)}(0),\; S^{(l)}(0) = w^{(l)}.\text{shape}$$

$$V^{(l)}(0),\; S^{(l)}(0) = 0$$

$$\beta_1 = 0.9 \qquad \beta_2 = 0.999$$

TAREA 3 :

Descomposición de Hankel

+

SNN+RMSprop + Adam

Funciones de Activación:

Función de Activación

- 1. ReLu:

$$f(x)=\begin{cases} x, & x>0 \\ 0, & x\le 0 \end{cases} \quad , x\in \Re^d$$

- 2. L-ReLu:

$$f(x)=\begin{cases} 0.01x, & x<0 \\ x, & x\ge 0 \end{cases} \quad , x\in \Re^d$$

- 3. ELU:

$$f(x)=\begin{cases} a\left(e^{x}-1\right), & x\le 0 \\ x, & x>0 \end{cases} \quad , x\in \Re^d$$

- 4. SELU:

$$f(x)=\lambda\times\begin{cases} a\left(e^{x}-1\right), & x\le 0 \\ x, & x>0 \end{cases}$$
$$x\in \Re^d,\ \lambda=1.0507\ ,\ a=1.6732$$

Función de Activación

- 5. C-ReLU: $\boxed{f(x) = \max(0, x) + \cos(x), \quad x \in \Re^d}$

- 6. S-ReLU: $\boxed{f(x) = \max(0, x) + \sin(x), \quad x \in \Re^d}$

- 7. Tangente Hiperbólica (tanh): $\boxed{f(x) = \dfrac{e^x - e^{-x}}{e^x + e^x}, \quad x \in \Re^d}$

Función de Activación

- 8. mTanh:

$$f(x) = a \times Tanh(bx), \quad x \in \Re^d$$
$$a = 1.7159, \quad b = \frac{2}{3}$$

- 9. Bipolar sigmoid:

$$f(x) = \frac{2}{1+e^{-x}} - 1, \quad x \in \Re^d$$

- 10. Softsign:

$$f(x) = \frac{x}{1+|x|}, \quad x \in \Re^d$$

Función de Activación

- 11.- Softplus: $$\boxed{f(x) = \log\left(1 + e^{x}\right), \quad x \in \Re^{d}}$$

- 12.- Sigmoid Linear Unit: $$\boxed{f(x) = \frac{x}{1 + e^{-x}}, \quad x \in \Re^{d}}$$

CONTINUARÁ….