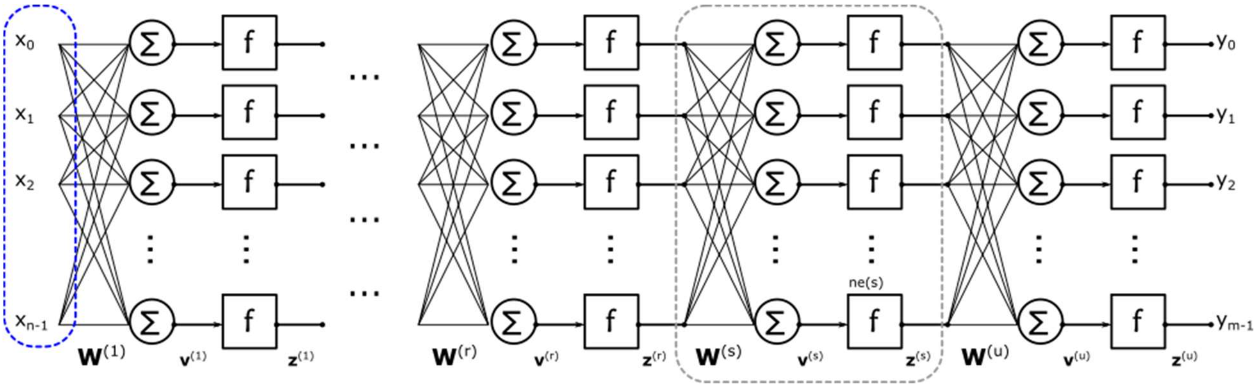
BPNN (Red Neuronal BackPropagation)

Partimos de {**X**, **Y**}, donde $\mathbf{X} \in \mathcal{M}_{p,n} \wedge \mathbf{Y} \in \mathcal{M}_{p,m}$, donde p se corresponde con el número de observaciones realizadas.

Supongamos una BPNN con una (1) capa de entrada, ene (n) capas de ocultas y una (1) capa de salida. Basaremos todo nuestro desarrollo en las tres últimas capas, las cuales serán identificadas como la capa (u) de salida, la capa (s) oculta, y cuyas activación se convierte en entrada de la capa (u), y por último la capa (r) también oculta, y cuya activación es entrada de la capa (s).



Las variables implicadas en nuestra red serían las siguientes:

- $\mathbf{x} \in \mathbb{R}^n$, vector de entrada a la red
- $\mathbf{W}^{(s)}$, matriz de pesos de la capa (s)
- $\mathbf{v}^{(s)}$, vector resultado de *net_input* de la capa (s)
- $\mathbf{z}^{(s)}$, vector resultado de la *activación* de la capa (s)
- $\mathbf{y} \in \mathbb{R}^m$, vector de salida de la red

Cada una de las capas de la red tiene asociada un número $ne^{(s)}$ de unidades de proceso, lo cual implica que $\mathbf{W}^{(s)} \in \mathcal{M}_{ne(s),ne(s-1)}$, $\mathbf{v}^{(s)} \in \mathbb{R}^{ne(s)}$ y $\mathbf{z}^{(s)} \in \mathbb{R}^{ne(s)}$.

Para el establecimiento de un algoritmo general de ajuste de los pesos, lo que hacemos es definir una Función de Energía E_q, que presente cierta relación con el error instantáneo que está cometiendo la red cuando predice una salida determinada *p*, $E_q = \frac{1}{2}\sum_{h=0}^{m-1}(\mathbf{y}_{p,h} - \mathbf{z}_{p,h}^{(u)})^2$, donde la capa (u) se corresponde con la capa de salida de la red, y donde ne(u) = m.

Con la definición anterior ya podemos establecer una forma de modificar las matrices de pesos $\mathbf{W}^{(s)}$ asociadas a todas las capas de la red: $\mathbf{W}^{(s)}(k+1) = \mathbf{W}^{(s)}(k) + \Delta\mathbf{W}^{(s)}$,

donde $\Delta w_{ji}^{(s)} = -\mu^{(s)}\nabla E_q = -\mu^{(s)}\frac{\partial E_q}{\partial w_{ji}^{(s)}}$.

Supongamos que estamos en la capa (u), la cual, y como hemos comentado anteriormente, se corresponde con la capa de salida de la red.

En este caso $\frac{\partial E_q}{\partial w_{ji}^{(u)}} = \frac{\partial}{\partial w_{ji}^{(u)}}\left[\frac{1}{2}\sum_{h=0}^{m-1}(\mathbf{y}_{p,h} - \mathbf{z}_{p,h}^{(u)})^2\right] = \frac{\partial}{\partial w_{ji}^{(u)}}\left[\frac{1}{2}\sum_{h=0}^{m-1}(\mathbf{y}_{p,h} - f(\mathbf{v}_h^{(u)}))^2\right]$.

Resolver la anterior derivada parcial se realiza teniendo en cuenta que $\frac{\partial E_q}{\partial w_{ji}^{(u)}} = \frac{\partial E_q}{\partial v_j^{(u)}}\frac{\partial v_j^{(u)}}{\partial w_{ji}^{(u)}}$, y por tanto, debemos resolver las dos derivadas parciales siguientes:

- $\frac{\partial E_q}{\partial v_j^{(u)}} = \frac{\partial}{\partial v_j^{(u)}}\left[\frac{1}{2}\sum_{h=0}^{m-1}(\mathbf{y}_{p,h} - f(\mathbf{v}_h^{(u)}))^2\right] = -(\mathbf{y}_{p,h} - f(\mathbf{v}_h^{(u)}))\,\dot{f}(\mathbf{v}_h^{(u)})$.
- $\frac{\partial v_j^{(u)}}{\partial w_{ji}^{(u)}} = \frac{\partial}{\partial w_{ji}^{(u)}}\left[\sum_{i=0}^{ne(s)-1} w_{ji}^{(u)} z_i^{(s)}\right] = z_i^{(s)}$, donde en este caso (s) indica la capa oculta (s) cuya activación $\mathbf{z}^{(s)}$ se convierte en entrada de la capa (u) de salida de la red.

Quedando la expresión de $\Delta\mathbf{W}^{(u)}$ para la capa de salida (u) de la siguiente forma:

$\Delta w_{ji}^{(u)} = \mu^{(u)}\left(\mathbf{y}_{p,h} - f\left(\mathbf{v}_h^{(u)}\right)\right)\dot{f}\left(\mathbf{v}_h^{(u)}\right) z_i^{(s)} = \mu^{(u)}\delta_j^{(u)} z_i^{(s)}$,

y donde $\delta_j^{(u)} = \left(\mathbf{y}_{p,h} - f\left(\mathbf{v}_h^{(u)}\right)\right)\dot{f}\left(\mathbf{v}_h^{(u)}\right)$, cuando estamos en la capa se salida.

Cuando estamos en una capa oculta genérica (s), incluida la capa (1) que recibe las entradas de la red, los cálculos asociados con la derivada parcial de la Función de Energía E_q se ven modificados de la siguiente forma:

$$\frac{\partial E_q}{\partial w_{ji}^{(s)}} = \frac{\partial E_q}{\partial v_j^{(s)}}\frac{\partial v_j^{(s)}}{\partial w_{ji}^{(u)}} = \frac{\partial E_q}{\partial z_j^{(s)}}\frac{\partial z_j^{(s)}}{\partial v_j^{(s)}}\frac{\partial v_j^{(s)}}{\partial w_{ji}^{(s)}}$$

Por tanto, debemos resolver las tres derivadas parciales anteriores, las cuales dan como resultado:

- $\frac{\partial E_q}{\partial z_j^{(s)}} = -\left[\sum_{h=0}^{m-1}\delta_h^{(u)} w_{hj}^{(u)}\right]$
- $\frac{\partial z_j^{(s)}}{\partial v_j^{(s)}} = \dot{f}(v_j^{(s)})$
- $\frac{\partial v_j^{(s)}}{\partial w_{ji}^{(s)}} = z_i^{(r)}$

Quedando la expresión de $\Delta\mathbf{W}^{(s)}$ para la capa oculta (s) de la siguiente forma:

$$\Delta w_{ji}^{(s)} = \mu^{(s)}\delta_j^{(s)} z_i^{(r)}$$

Donde en este caso $\delta_j^{(s)} = \left[\sum_{h=0}^{m-1}\delta_h^{(u)} w_{hj}^{(u)}\right]\dot{f}\left(\mathbf{v}_j^{(s)}\right)$.