Práctica de Redes Neuronales Recurrentes

Esther Cuervo Fernández

11 de enero de 2018

Parte I
Red de Elman

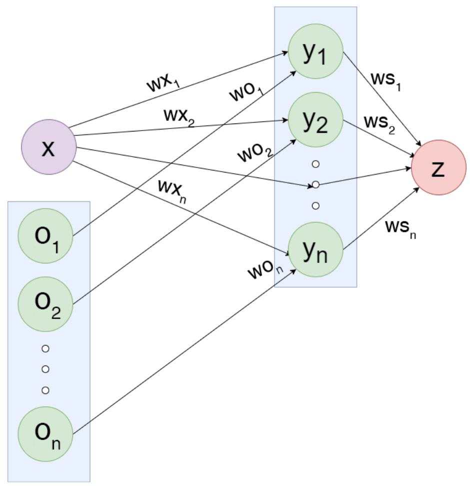
Una Red de Elman se trata de una red neuronal con la particularidad de que los outputs de la capa oculta para el instante de tiempo $t - 1$ son introducidos como input de la red en el momento t.

Utilizaremos esta red para hacer predicciones sobre los valores en bolsa de las acciones de Iberdrola entre las fechas `28/Diciembre/2016` y `1/Diciembre/2017` para un total de 12 meses de 20 días cada uno.

1. Arquitectura de la red

La red contendrá una capa de entrada, otra oculta, y otra de salida, más una capa auxiliar que contiene las salidas de la capa oculta para el instante anterior.

Las conexiones son las siguientes:



Siendo x la salida de la neurona de la capa de entrada, z la salida de la neurona de la capa de salida, y_i las salidas de las neuronas de la capa oculta en el instante actual, y o_i las salidas de las neuronas de la capa oculta en el instante anterior.

Cada conexión tiene un peso, designado como wx_j para las conexiones entre la neurona de entrada y la neurona j de la capa oculta, ws_j para las conexiones entre la neurona de la capa oculta j y la neurona de salida.

Los pesos wo_{ij} conectan la neurona de la capa auxiliar o_j con la neurona de la capa oculta y_i.

Además cada neurona y_i de la capa oculta tiene una conexión extra, conocida como `término bias`, con valor de entrada siempre igual a 1, y con peso wxb_j para la neurona de la capa oculta j. La neurona de salida también tiene una conexión bias con peso wsb.

2. Preprocesado

Los datos a utilizar se encuentran en el documento `originalFiles/dat_entrada.csv`, formateado como el valor de cada día separados por comas. Contamos con 12 meses de 20 días cada uno, por lo que el fichero contiene 240 valores.

Para entrenar nuestra red Elman realizaremos un Hold-Out no aleatorio a nuestros datos, tal que utilizamos $\frac{2}{3}$ de las secuencias para entrenamiento, y $\frac{1}{3}$ para validación.

Definimos secuencia como un conjunto de 21 valores continuos en el tiempo, tal que la secuencia que comienza en el momento t será:

$$X(t) = \{x(t), x(t+1), x(t+2), ..., x(t+19), x(t+20)\}$$

El último valor de cada secuencia, $x(t+20)$ será la salida deseada de la red para esa secuencia.

Por tanto nuestros datos cuentan con $240 - 20 = 220$ secuencias:

$$X(1) = \{x(1), x(2), ..., x(21)\}$$

$$...$$

$$X(220) = \{x(220), x(221), ..., x(240)\}$$

Por tanto seleccionamos las primeras $\frac{220*2}{3} \approx 147$ secuencias para entrenamiento, y las siguientes 73 secuencias para validación.

Se necesita realizar un escalado de ambos conjuntos de datos, pero esto se realiza por cada secuencia, ya que son estos 20 datos los que afectan a la predicción.

La normalización sigue la siguiente fórmula:

$$x_i = \frac{x_i - (min(x) - min(x) * 0{,}2)}{(max(x) + max(x) * 0{,}2) - (min(x) - min(x) * 0{,}2)} \quad \forall i$$

Siendo x la secuencia correspondiente y x_i el valor i de dicho secuencia.

3. Entrenamiento

Tras esto comenzamos con el entrenamiento de la red, que iterará sobre todas las instancias de entrenamiento un número de épocas pre-definido, y realizará 20 fases hacia delante y una fase hacia atrás para cada secuencia.

3.1. Fase hacia delante

Neurona de entrada La neurona de entrada introduce la secuencia valor a valor, por tanto en el instante t de la fase hacia delante de la secuencia que comienza en el día i la salida de la neurona de entrada es:

$$x^t = x(i + t)$$

Retroalimentación Las neuronas de la capa auxiliar, o, tendrán como salidas las salidas en el instante $t-1$ de la capa oculta. Si estamos en el instante $t = 1$, estas neuronas tienen salida 0.

$$o_i^1 = 0$$
$$o_i^t = y_i^{t-1}$$

Neuronas de la capa oculta Las neuronas de la capa oculta, y_i reciben como entrada x^t y o^t, y su salida es:

$$y_i^t = F(wx_i * x^t + wxb_i + \sum_{j=0}^{N[O]} wo_{ij} * o_j^t)$$

Siendo $N[O]$ el número de neuronas auxiliares.

Neurona salida La neurona de la capa de salida z tendrá como entrada cada salida de la capa oculta, y^t y su salida es:

$$z^t = F(\sum_{j=1}^{N[H]} (ws_j * y_j^t) + wsb)$$

Con $N[H]$ igual al número de neuronas en la capa oculta.

Cabe destacar que la salida z^t solo será de iportancia en la red cuando lleguemos a $t = 20$, en cuyo momento la utilizaremos como salida obtenida para compararla con el último valor de la secuencia.

3.2. Fase hacia atrás

Durante esta fase realizamos una retropropagación del error, modificando los pesos de las conexiones entre neuronas para intentar mejorar la capacidad de predicción de la red. Esto solo se realiza al finalizar la fase hacia delante en el instante $t = 20$.

En esta fase se utiliza una constante γ, el *factor de aprendizaje*.

Neurona de salida La ecuación del error es la siguiente:

$$E = \frac{1}{2}(z^{20} - d)^2$$

Siendo d la salida deseada, es decir, el último valor de la secuencia actual, para la secuencia que comienza en el día i este valor es x(i+20) en el dataset.

Por tanto el cambio del peso para cada i de las $N[H]$ conexiones con las neuronas de la capa oculta es:

$$\begin{aligned}
\Delta ws_i &= \gamma\left(\frac{-\partial E}{\partial ws_i}\right)\\
\frac{-\partial E}{\partial ws_i} &= \frac{-\partial E}{\partial z^{20}} * \frac{\partial z^{20}}{\partial ws_i}\\
\frac{\partial E}{\partial z^{20}} &= 2\frac{1}{2}(z^{20} - d) = (z^{20} - d)\\
\frac{-\partial E}{\partial z^{20}} &= -(z^{20} - d) = (d - z^{20})\\
\frac{\partial z^{20}}{\partial ws_i} &= \frac{\partial F(u^s)}{\partial ws_i}
\end{aligned}$$

$$\begin{aligned}
u^s &= \sum_{j=1}^{N[H]}(ws_j * y_j^{20}) + wsb\\
\frac{\partial F(u^s)}{\partial ws_i} &= F'(u^s) * y_i^{20}
\end{aligned}$$

$$\Delta ws_i = \gamma * (d - z^{20}) * F'(u^s) * y_i^{20}$$

F' es la derivada de la función de activación utilizada en la capa de salida durante la fase hacia delante.

Esta fórmula cambia ligeramente para el cálculo del incremento del peso del *bias*, debido a que este peso va multiplicado por 1 en u^s:

$$\Delta wsb = \gamma * (d - z^{20}) * F'(u^s)$$

Neuronas de la capa oculta Para las conexiones que van desde la neurona de entrada a la neurona de la capa oculta i las ecuaciones del incremento de pesos son:

$$\begin{aligned}
\Delta wx_i &= \gamma \left(\frac{-\partial E}{\partial wx_i} \right) \\
\frac{-\partial E}{\partial wx_i} &= \frac{-\partial E}{\partial y_i^{20}} * \frac{\partial y_i^{20}}{\partial wx_i} \\
\frac{-\partial E}{\partial y_i^{20}} &= \frac{-\partial E}{\partial z^{20}} * \frac{\partial z^{20}}{\partial y_i^{20}} \\
\frac{-\partial E}{\partial y_i^{20}} &= (d - z^{20}) \frac{\partial z^{20}}{\partial y_i^{20}} \\
\frac{\partial z^{20}}{\partial y_i^{20}} &= \frac{\partial F(u^s)}{\partial y_i^{20}} = F'(u^s) * ws_i
\end{aligned}$$

$$\begin{aligned}
u_i^h &= wx_i * x^{20} + wxb_i + \sum_{j=0}^{N[O]} wo_{ij} * o_j^{20} \\
\frac{\partial y_i^{20}}{\partial wx_i} &= \frac{\partial F(u_i^h)}{\partial wx_i} = F'(u_i^h) * x^{20}
\end{aligned}$$

$$\Delta wx_i = \gamma * (d - z^{20}) * F'(u^s) * ws_i * F'(u_i^h) * x^{20}$$

Para la conexión que va desde la neurona auxiliar o_j a la entrada de la neurona oculta y_i las ecuaciones cambian para el último término:

$$\begin{aligned}
\Delta wo_{ij} &= \gamma (\frac{-\partial E}{\partial wo_{ij}}) \\
\frac{-\partial E}{\partial wo_{ij}} &= \frac{-\partial E}{\partial y_i^{20}} * \frac{\partial y_i^{20}}{\partial wo_{ij}} \\
&\dots \\
\frac{\partial y_i^{20}}{\partial wo_{ij}} &= \frac{\partial F(u_i^h)}{\partial wo_{ij}} = F'(u_i^h) * o_j^{20}
\end{aligned}$$

$$\Delta wo_{ij} = \gamma * (d - z^{20}) * F'(u^s) * ws_i * F'(u_i^h) * o_j^{20}$$

De igual manera la conexión del *bias* de la neurona oculta i tiene el siguiente incremento de peso:

$$\Delta wxb = \gamma * (d - z^{20}) * F'(u^s) * ws_i * F'(u_i^h)$$

Término momento Además de las fórmulas del incremento del error indicadas anteriormente, se añade un término momento, que es igual al incremento de peso que ha experimentado el peso para la secuencia anterior. Este término trata de evitar que la red caiga en mínimos locales. Irá multiplicado por un factor de inercia α.

Actualización de pesos Por tanto los nuevos pesos tras la secuencia r son:

$$\begin{aligned}
wx_i^r &+= \Delta^r wx_i + \alpha(\Delta^{r-1} wx_i) \\
ws_i^r &+= \Delta^r ws_i + \alpha(\Delta^{r-1} ws_i) \\
wo_{ij}^r &+= \Delta^r wo_{ij} + \alpha(\Delta^{r-1} wo_{ij}) \\
wxb^r &+= \Delta^r wxb + \alpha(\Delta^{r-1} wxb) \\
wsb^r &+= \Delta^r wsb + \alpha(\Delta^{r-1} wsb)
\end{aligned}$$

4. Resultados

Parte II

Red de Jordan