

Alumno 1

Pregunta.-

-Con respecto al tema de Redes, no llego a entender bien las diapositivas 28, 29 y 30.

Respuesta.-

Las tres transparencias están relacionadas con la hipótesis de que cualquier función real continua sobre D , un conjunto compacto en \mathbb{R}^n , se puede representar o ajustar mediante una red neuronal con activación hacia adelante y formada por una capa de entrada constituida por las características o variables independientes del modelo, una capa oculta formada, en el caso de las redes de tipo perceptron multicapa, por funciones sigmodales, y una capa de salida constituida por las variables dependientes del problema.

En la transparencia 28 está expuesta la demostración de Chen de 1991, en la 29 se muestra algunas de las variantes del teorema anterior, donde en la columna 1 se muestra el autor de la hipótesis a demostrar, en la columna 2, las condiciones que tiene que tener la función de activación de la red neuronal, en la columna 3 el dominio donde tiene que estar definida la función a justar y en la columna 4, el tipo de método de demostración del teorema.

En la transparencia 30 se muestra la forma práctica en la que vamos a utilizar los modelos de redes neuronales con activación hacia adelante

Estas transparencias son adicionales y no vamos a entrar en ellas, salvo la 30, donde las redes con activación hacia adelante se van a desarrollar con posterioridad. Están puestas para que vean que existe una teoría matemática muy fuerte que hace que, con ciertas restricciones en las funciones a justar, las redes neuronales sean reconocedores universales, y la cuestión es determinar el nº de nodos en capa oculta para obtener un modelo que se ajuste a la función dada.

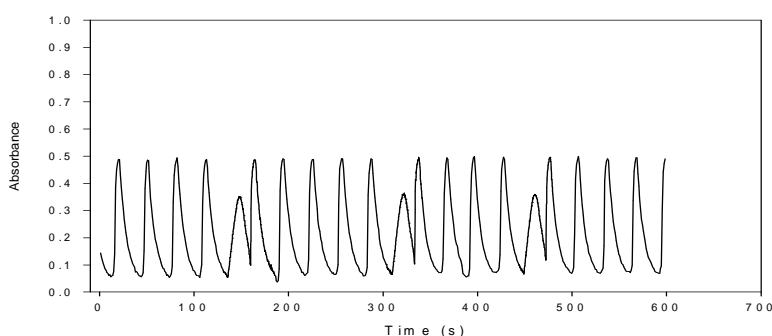
Pregunta.-

-En la diapositiva 45 no sé de dónde salen esos números ni entiendo bien lo que se está calculando por el tema de "Redes Neuronales Podadas".

Respuesta.-

En la transparencia 44 se muestra un grafo de una red neuronal entrenada para un problema de mezclas de compuestos químicos, donde se conocen los parámetros fundamentales de las curvas.

El problema está asociado a la resolución de mezclas de especies basado en los efectos de la perturbación en las reacciones químicas oscilantes, lo que implica el desarrollo de un procedimiento de dos pasos para obtener modelos no lineales para predecir la concentración de los componentes en tales mezclas. El enfoque se probó en la determinación simultánea de dos derivados fenólicos relacionados, como el pirogalol (P) y el ácido gálico (GA), a través de sus efectos de perturbación en la reacción química oscilante clásica que implica la oxidación del ácido malónico por el ion bromato en un sulfúrico medio ácido catalizado por sales de cerio (IV), que se conoce como la reacción de Belousov-Zhobotinskii (ver figura 1). Esta reacción exhibe cambios periódicos en la concentración de algunas especies que se reflejan en cambios potenciales o en el color cíclico, el último directamente relacionado con el ciclo catalítico, que implica cambios de color entre amarillo [cerio (IV)] e incoloro [cerio (III)] (ver oscilaciones regulares en la Figura 1)



Se ha observado que el conjunto de señales (t_i , S_{ii}) se ajustó con precisión por regresión de mínimos cuadrados a una curva gaussiana.

Así, el modelo no lineal viene dado por

$$S_{t_i} = a_m e^{\frac{1}{2} \frac{(t_i - t_m)^2}{s^2}} + \varepsilon_i$$

El principio de mínimos cuadrados se usó para estimar los parámetros mediante el método de Levenberg-Marquardt. Las estimaciones obtenidas \hat{a}_m , \hat{s} , y \hat{t}_m se usan como entradas a la red neuronal, de tal manera que todos los modelos de redes neuronales propuestos tienen tres variables en la capa de entrada.

Un segundo paso implica un diseño de minimización de la red neuronal mediante la eliminación sucesiva de pesos después de que la red ha sido entrenada con un rendimiento satisfactorio. Este proceso de aprendizaje se lleva a cabo utilizando el algoritmo de aprendizaje de retropropagación del error EDBD, 9 donde la función objetivo de poda de conexiones o pesos del modelo es independiente de la función objetivo de entrenamiento del modelo, la cual al ser un problema de regresión es el error cuadrático medio.

Así, la función de regularización es de la forma

$$\lambda n_w \log \sum_{k=1}^{n_w} |w_k|$$

esta función representa la complejidad de la red en función del valor absoluto de las magnitudes de peso, w_k , donde n_w es el número de pesos de la red neuronal. La función evaluada se basa en la hipótesis de que la distribución "a priori" de los pesos de la red sigue una distribución estadística laplaciana. El parámetro λ es un término de regularización, que representa la importancia relativa del término de penalización de la complejidad de la red con respecto al término de medida del error cometido en la predicción tanto del pirogalol (P) como del ácido gálico (GA), definida como:

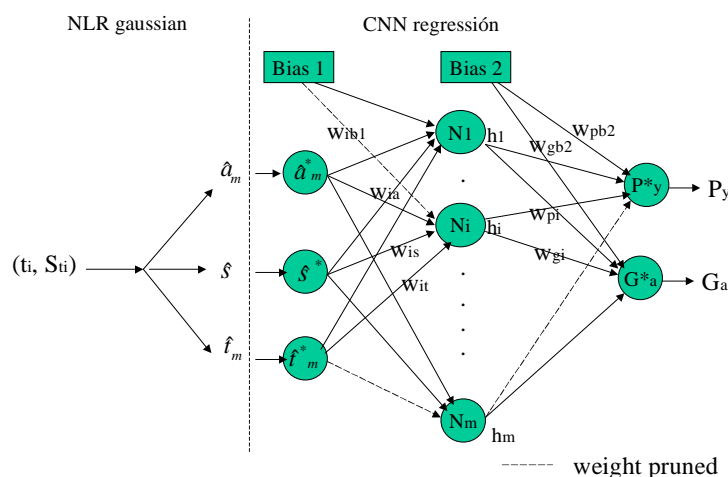
$$\frac{1}{2} n_T \log \sum_{p=1}^{n_T} (y_p - o_p)^2$$

Este término es la suma de los errores al cuadrado entre los valores de salida reales (o_p , valores conocidos de pirogalol (P) como del ácido gálico (GA) de las señales de mezcla) y los valores de salida objetivo (y_p , valores que nos devuelve la red neuronal entrenada), y n_T es el número de patrones del conjunto de entrenamiento. Se utilizaron funciones sigmodes y lineales para los nodos de la capa oculta y nodos de la capa de salida, respectivamente. Además, para evitar problemas de saturación cuando se utilizan funciones sigmoides, (estas funciones tienen asíntotas en $y=0$ y en $y=1$), los valores de los nodos de entrada y salida se normalizaron en el rango de 0.1 a 0.9.

Por lo tanto, los valores de entrada normalizados son, \hat{a}_m^* , \hat{s}^* y \hat{t}_m^* , y para los nodos de salida correspondientes a la concentración de pirogalol, es $[P]^*$ y para el ácido gálico, $[GA]^*$. En la Figura 2 podemos ver el modelo de red neuronal.

La precisión de cada modelo se evaluó a partir de los resultados obtenidos para el conjuntode test utilizando el error estándar de predicción relativo (SEP)

$$SEP = \frac{100}{A_i} \sqrt{\frac{\sum_{i=1}^n (A_i - \hat{A}_i)^2}{n}}$$



Procedimiento de estimación de los parámetros del modelo

donde A_i y \hat{A}_i se encuentran los valores esperados para la concentración de analito en la mezcla, \bar{A}_i es su valor promedio y n es el número de patrones del conjunto de test.

Como se puede ver en la Tabla 2, los modelos con mejor valor de SEP y un menor número de conexiones fueron los que tienen una topología de red 3:3:2 (tres variables de entrada, 3 nodos sigmoides en capa oculta y 2 nodos en capa de salida) y una razon del conjunto de datos de entrenamiento/test de 44/22, y una topología de red 3:4:2 con una proporción del conjunto de datos 54/27 entrenamiento/test. Las ecuaciones de los modelos propuestos se pueden ver en la Tabla 3

Tabla 2. Precisión del algoritmo (SEP en entrenamiento y test) utilizando diversas topologías de red y tamaños de conjuntos de datos aplicados a la resolución de mezclas de pirogalol y ácido gálico

topología de red	data set train/test	Pyrogallol					
		Media			Desviación típica		
		SEP train	SEP test	Nº Conn.	SEP train	SEP test	Nº Conn.
3:3:1	44/22	5.07	4.83	11.5	0.60	0.34	2.83
3:2:1	54/27	5.86	5.64	9.0	0.46	0.44	1.41
3:4:2	44/22	4.73	4.39	21.7	0.27	0.41	3.19
		mejor			peor		
		SEP train	SEP test	Nº Conn	SEP train	SEP test	Nº Conn
3:3:1	44/22	3.79	4.19	14	5.67	5.19	14
3:2:1	54/27	4.76	4.65	8	6.22	6.12	9
3:4:2	44/22	5.22	4.08	19	4.83	5.27	22
3:3:2	54/27	5.57	4.40	18	5.97	5.17	19
		Acido Gálico					
		Media			Desviación típica		
		SEP train	SEP test	Nº Conn.	SEP train	SEP test	Nº Conn.
3:2:1	44/22	5.44	5.66	9.7	0.36	0.17	0.67
3:3:1	54/27	10.80	10.01	11.1	0.44	0.85	1.91
3:4:2	44/22	5.13	6.01	21.7	0.29	0.30	3.19
3:3:2	54/27	9.78	9.52	18.2	0.21	0.38	0.91
		mejor			peor		
		SEP train	SEP test	Nº Conn	SEP train	SEP test	Nº Conn
3:2:1	44/22	5.50	5.38	9	5.89	6.01	10
3:3:1	54/27	10.01	8.51	15	11.37	11.07	10
3:4:2	44/22	4.92	5.70	25	5.21	6.65	24
3:3:2	54/27	9.67	8.99	16	10.09	10.24	17

Tabla 3. Ecuaciones obtenidas por los modelos de redes neuronales de tipo MLP podadas aplicadas a la resolución de mezclas de pirogallol y ácido gálico

modelo 5 (NT: 3:4:2; C: 19; DS: 44/22)	modelo 6 (NT: 3:3:2; C: 18; DS: 54:27)
$[P]^* = 1.25h_1 - 0.35h_2 + 0.09h_4$	$[P]^* = 1.54h_1 + 0.37h_2 - 0.78h_3$
$[GA]^* = -0.19 - 0.25h_1 + 1.74h_2 - 0.12h_3 - 0.34h_4$	$[GA]^* = -0.69 - 1.74h_1 + 1.89h_2 + 1.22h_3$
$h_1 = \frac{1}{1 + \exp(2.45 - 2.18 \hat{a}_m^* - 1.82 \hat{s}^* - 0.93 \hat{t}_m^*)}$	$h_1 = \frac{1}{1 + \exp(3.69 - 0.07 \hat{a}_m^* - 3.54 \hat{s}^*)}$
$h_2 = \frac{1}{1 + \exp(0.59 - 3.6 \hat{a}_m^* + 2.07 \hat{s}^*)}$	$h_2 = \frac{1}{1 + \exp(0.74 - 3.71 \hat{a}_m^* + 1.5 \hat{s}^* - 0.38 \hat{t}_m^*)}$
$h_3 = \frac{1}{1 + \exp(0.79)}$	$h_3 = \frac{1}{1 + \exp(0.8 - 1.93 \hat{a}_m^* + 7.05 \hat{s}^* + 2.3 \hat{t}_m^*)}$
$h_4 = \frac{1}{1 + \exp(3.56 \hat{a}_m^* - 9.35 \hat{s}^* - 4.41 \hat{t}_m^*)}$	

NT: topología de red; C: n° conexiones; DS: razón de los conjuntos de datos (entrenamiento/test); hi: salida del nodo i-ésimo de la capa oculta