

Tarea 5

1)

RBF

Se implementó una red de base radial utilizando el lenguaje de programación Python para interpolar los datos encontrados en el archivo rabbit.csv. El conjunto de entrenamiento tiene la forma $\{(X_i, Y_i)\}$ con $i=1$ hasta N , y tenemos un conjunto de centros $\{T_i\}$ con $i=1$ hasta K , $N \geq K$. La red consta de una capa de oculta, cuyos pesos no son valores fijos, sino una función de distancia entre el dato a ser procesado y cada uno de los centros T_i escogidos (uno por neurona). La función de activación de cada neurona es una función de activación radial $F(r, \sigma) = e^{-\frac{r^2}{2\sigma^2}}$ con valores de sigma mayores a cero y r perteneciente a los números reales. Luego de la capa de la oculta se encuentra la capa de salida, con función de activación lineal, y cuyos pesos son determinados por el entrenamiento de la red resolviendo el sistema de ecuaciones $\Phi w = D$, donde D es el vector de valores esperados para cada dato del conjunto de entrenamiento, w es el vector de pesos de la neurona de salida, y Φ es una matriz de dimensiones $N \times K$, y $\Phi_{ij} = F(X_i, T_j) = e^{-\frac{(X_i - T_j)^2}{2\sigma^2}}$. Se realizó el enteramiento sin procesar los datos, para diferentes cantidades de centros y valores de dispersión σ , y no se obtuvieron resultados satisfactorios. Se procedió a entrenar procesando el eje Y, una vez más sin obtener buenos resultados. Finalmente se procedió a pre procesar el eje X utilizando la técnica de mínimo y maximo, donde sí se lograron los resultados esperados, y donde buscamos la mejor topología de la red.

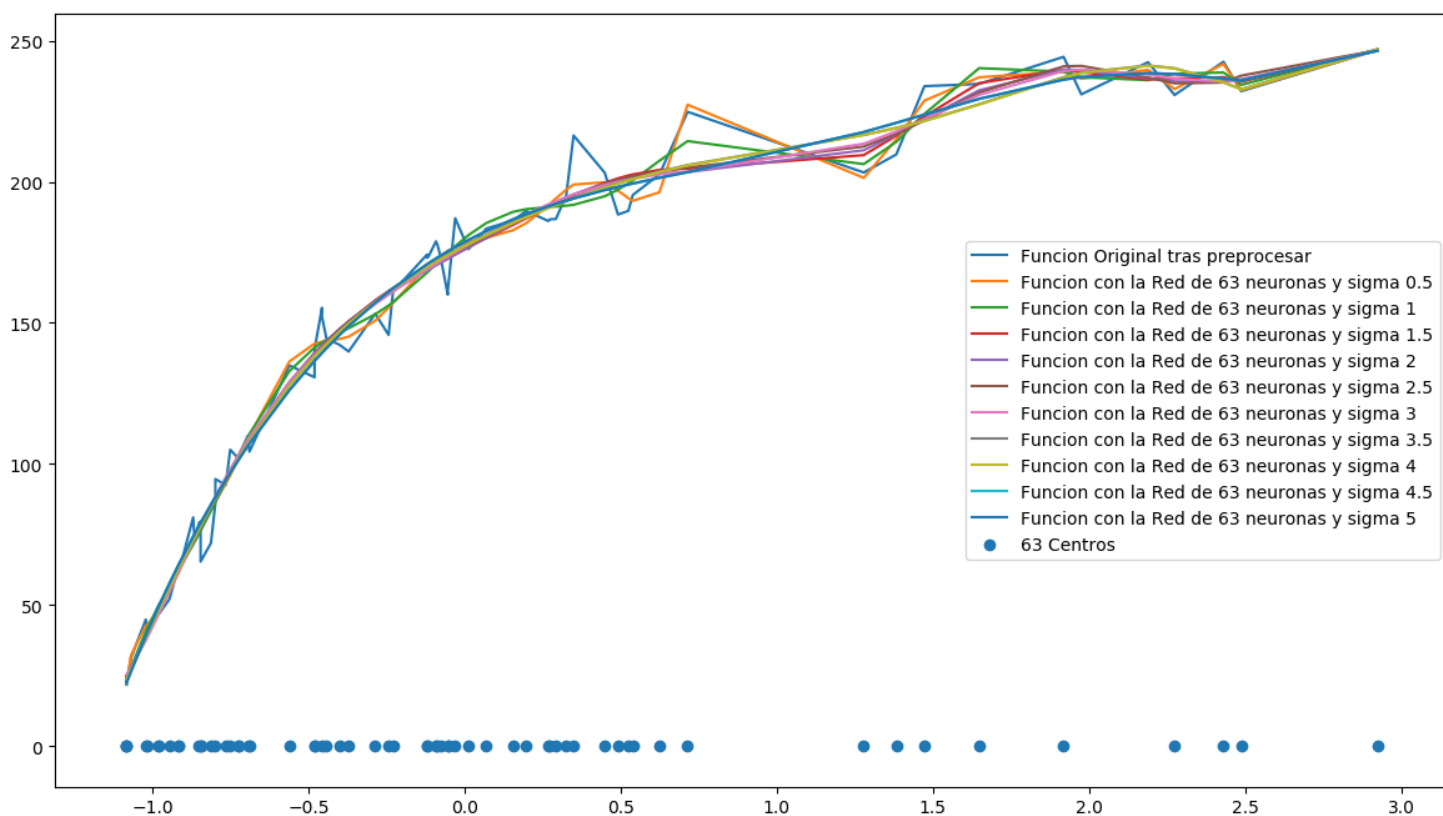
Se realizó la búsqueda con una combinación de valores de sigmas y de porcentaje de centros con respecto a la cantidad de datos, y se obtuvieron los siguientes resultados:

%Centros\Sigmas	0,5	1	1,5	2	2,5
90	34,0192085	46,9476753	54,1855707	54,5851446	54,8893515
75	35,9314243	47,1416265	54,1831803	54,5941819	55,0934204
50	37,1811516	47,6880399	54,4078653	54,7773035	55,3921854
25	47,6991589	53,4158581	54,6515747	55,4846341	56,7096767

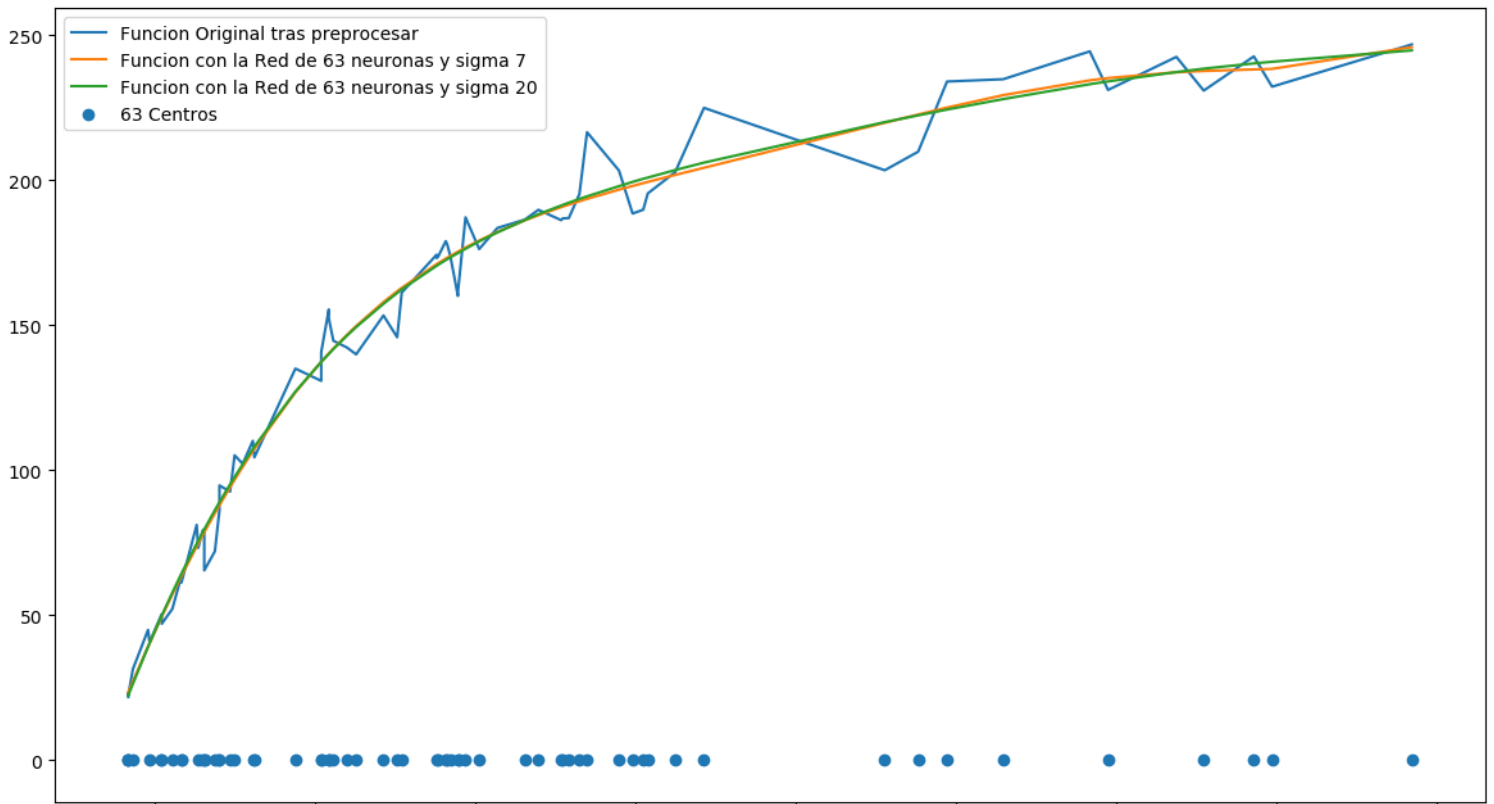
%Centros\Sigmas	3	3,5	4	4,5	5
90	55,4278373	57,0389136	57,0903690	58,7322993	58,7339986
75	55,4353972	57,0419663	57,0702808	58,7325500	58,7360300
50	55,4855925	56,9870681	59,3496659	58,7625492	58,7581727
25	56,8597213	58,8203221	58,7994997	58,9072567	58,7730846

Error cuadrático medio según el porcentaje de centros tomados y el valor de dispersión sigma usado

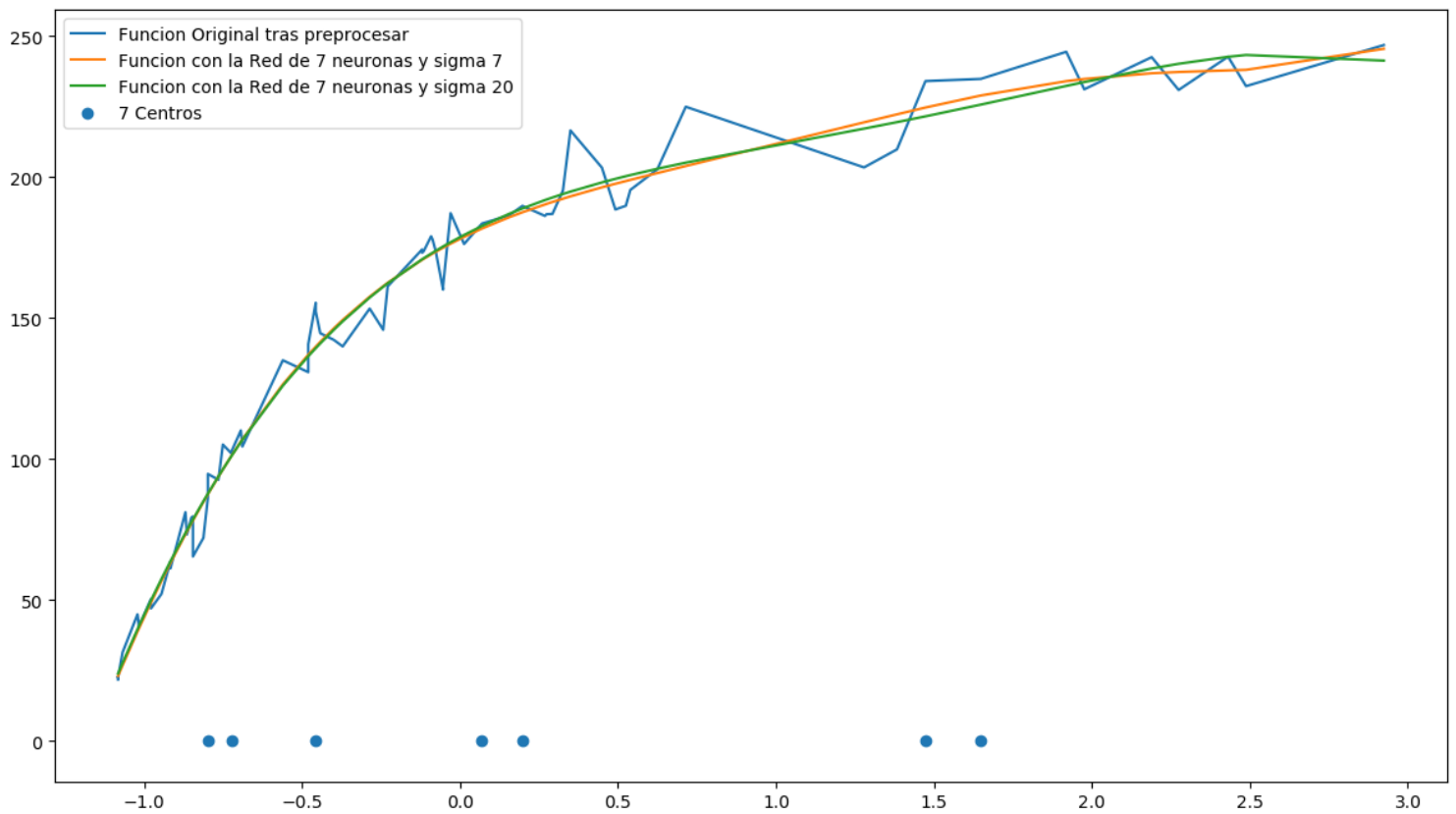
Los mejores resultados se observan cuando se tienen mayor cantidad de neuronas y sigmas más bajos, sin embargo esto puede ser signo de que se está incurriendo en overfitting y ajustando demasiado la red a los valores de entrenamiento. Esto se observa especialmente en las gráficas de las funciones obtenidas, tomando en cuenta la poca cantidad de datos de entrenamiento y el ruido que se observa en los datos finales de la función. Para mejorar la generalización, tendemos a preferir a la función “más suave”, se observa que al tener valores de sigma altos, la función obtenida se suaviza considerablemente. Realizamos varias comparaciones, resultando en una comparación final con valores de sigma 7 y 20, en las cuales no se encuentra mayor diferencia, prefiriendo entonces sigma 7. Probamos ahora reducir el número de neuronas, y obtenemos resultados similares, por lo que nos quedamos con una red de 7 neuronas, ya que se prefieren redes mas simples que realicen el trabajo correctamente.



Funciones obtenidas por la red con 90% de los datos usados como centros.

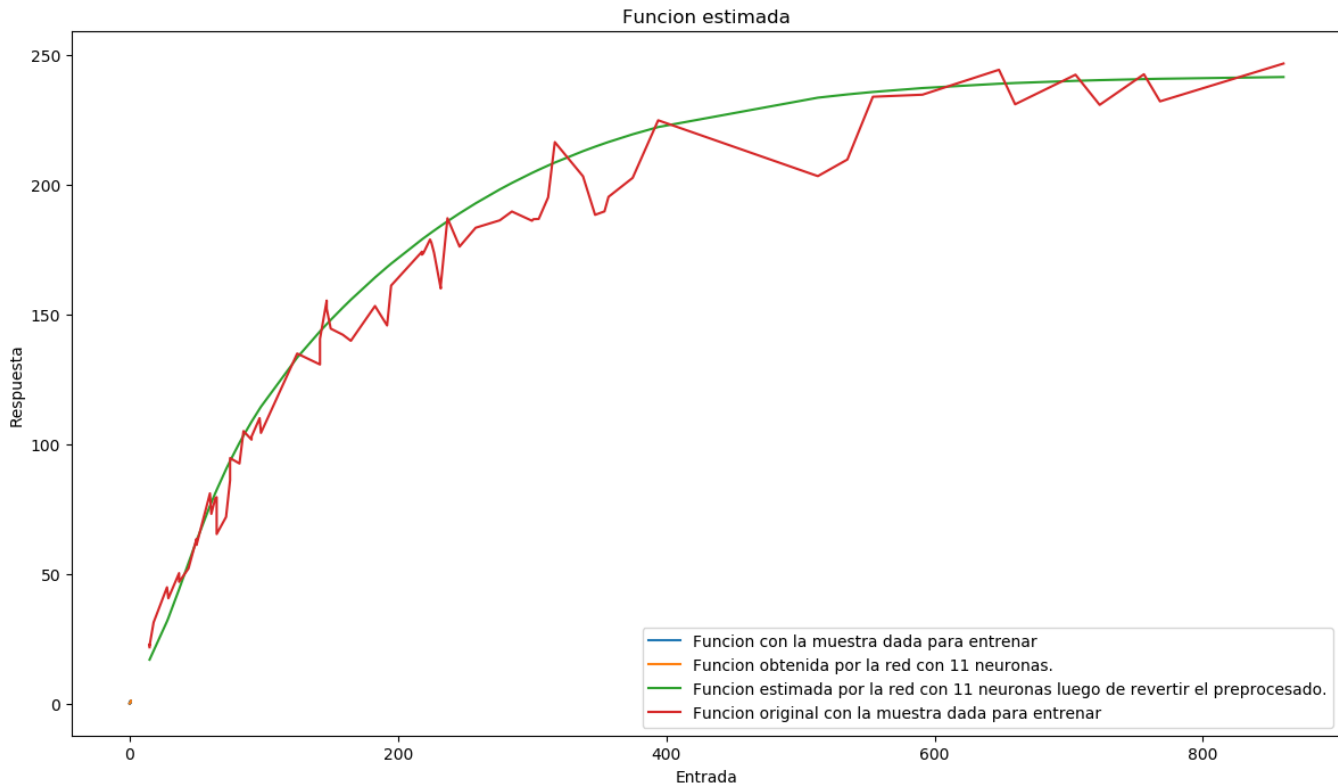


Comparación con el uso de sigma 7 y 20



Mejor Red encontrada: 7 Neuronas y Sigma 7

Realizando una comparación de esta RBF con la MLP usada para resolver el mismo problema, y observando las gráficas obtenidas por ambas redes, se puede deducir que la RBF tiene mejor capacidad de generalización y realiza un mejor trabajo que la MLP, ya que esta última tiende a arrojar resultados un poco por encima de los esperados. Cabe mencionar que la RBF aun realizando un mejor trabajo que la MLP, es una red considerablemente más sencilla de realizar que las MLP.



Funcion obtenida por la MLP para el mismo problema.

2) Se implementó una red de base radial para interpolar la función $h(x)=0,5+0,4\sin(2\pi x)$, se tomaron 100 puntos de forma aleatoria de una distribución uniforme en el intervalo $[0,1]$, se usaron estos puntos para evaluar la función, y cada resultado se le agregó un ruido tomado aleatoriamente de una distribución normal con media 0 y varianza 0.1 (se toma esta varianza dada la magnitud de los valores resultantes). Usando este conjunto de datos con ruido, realizamos el entrenamiento de la red bajo la misma metodología del problema anterior. Para este caso, como contamos con los valores exactos que debe dar la función, podemos decir la mejor topología de la red directamente observando el error cuadrático medio producido por cada red. Se realizó el entrenamiento de distintas redes con distintos valores de dispersión, los resultados se muestran en la siguiente tabla:

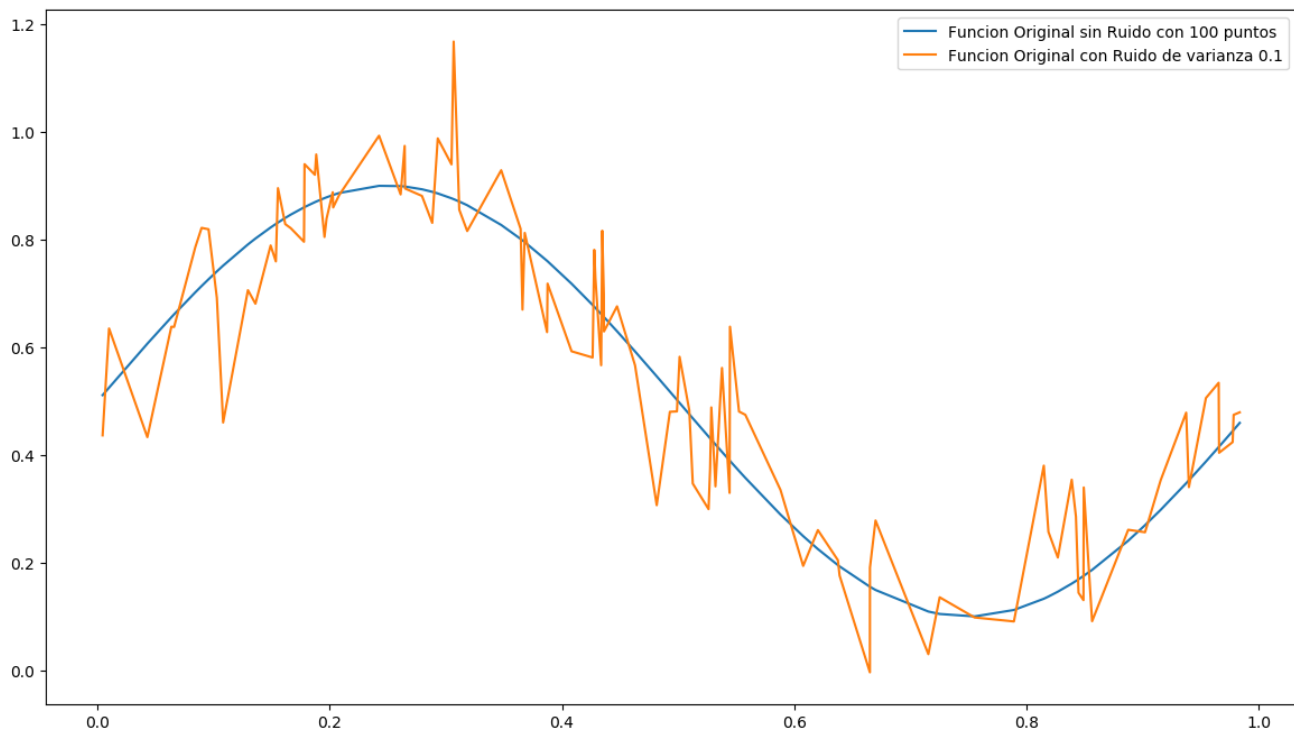
%Centros\Sigma	0.5	1	2.5	5
100	0.00081098	0.00045831	0.00030674	0.00029242
90	0.00080965	0.00045159	0.00034186	0.00081091
75	0.0008123	0.00046719	0.00033977	0.00030516
50	0.00080338	0.00045711	0.00030754	0.00028309
20	0.00084179	0.00047088	0.00036502	0.00081087

%Centros\Sigma	7	10	15	20
100	0.00081611	0.00082135	0.0008379	0.00083824
90	0.000816	0.00081707	0.00083792	0.00083826
75	0.00081635	0.0008183	0.00083794	0.00083881
50	0.00081598	0.00082269	0.00083778	0.00083825
20	0.00081648	0.00083981	0.00083794	0.00083844

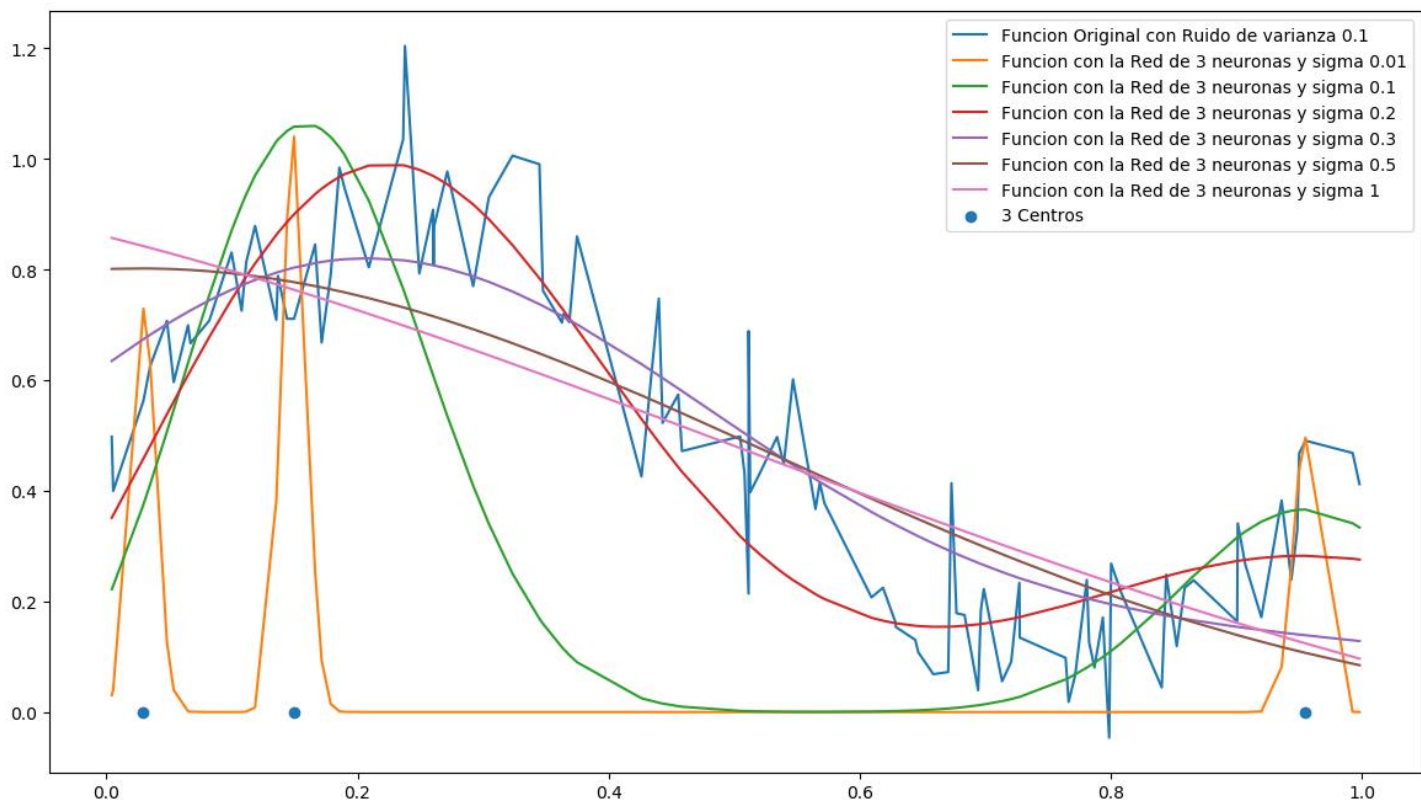
% Centros \ Sigmas	0.01	0.1	0.2	0.3	0.5	1
3	0.26641544	0.09453633	0.00938965	0.01279038	0.02426057	0.02893491
4	0.2610894	0.01441994	0.00683346	0.0018068	0.00026817	0.0005036
5	0.29589508	0.15112545	0.00625816	0.00037434	0.00016735	0.00056299
7	0.3011625	0.14358023	0.00086068	0.00020923	0.00023548	0.00023589
8	0.2449152	0.02694281	0.00017144	0.00023277	0.00023576	0.00023493
9	0.22572136	0.00110722	0.00040895	0.00026992	0.0002361	0.00023959
10	0.2119545	0.00129423	0.00088444	0.00088011	0.00086693	0.00028156

Error cuadrático medio según el porcentaje de centros tomados y el valor de dispersión sigma usado

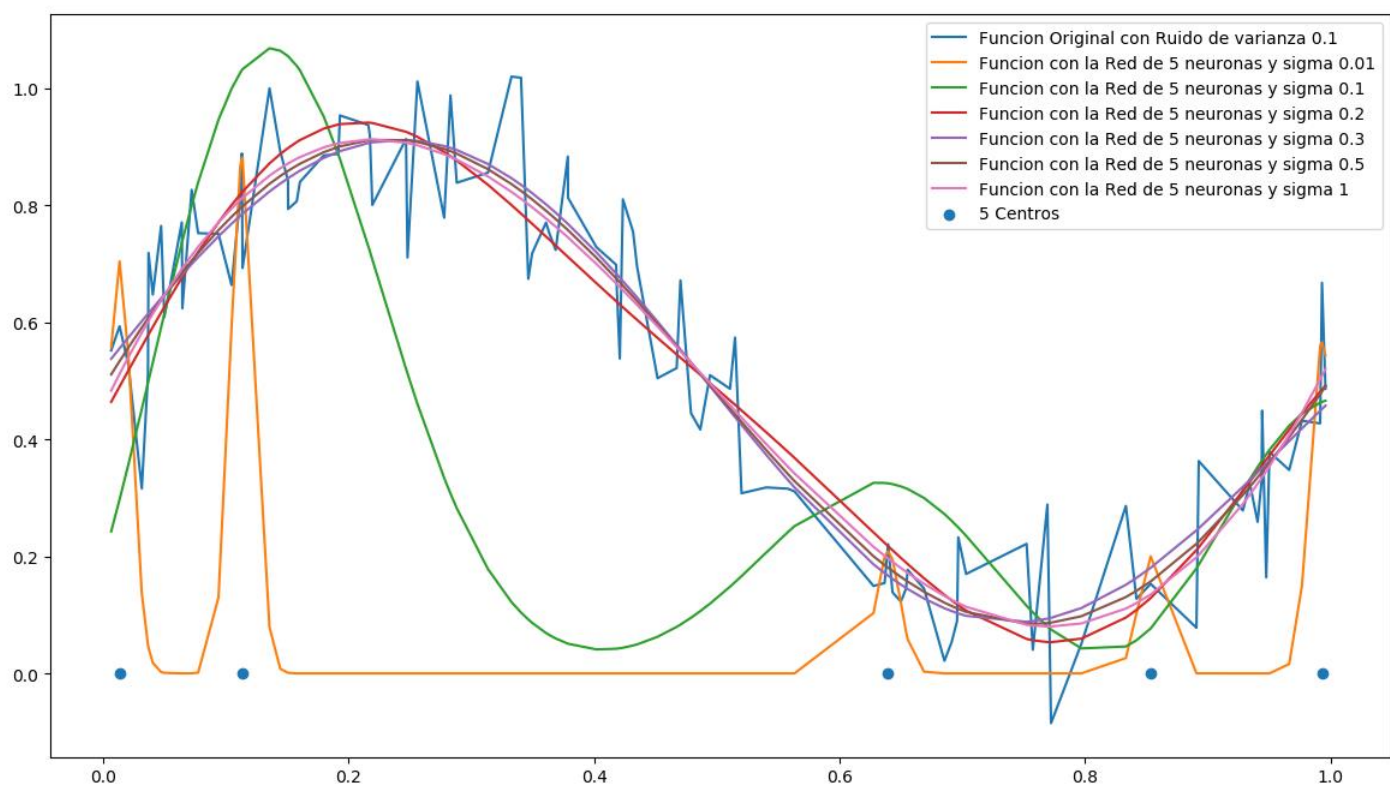
Observando la tabla podemos notar que se necesitan al menos 4 neuronas para realizar un buen trabajo, y que el parámetro de dispersión optimo se encuentra entre 0.3 y 0.5. En base a esto definimos que la mejor topología de red para este problema es de 5 neuronas con parámetro de dispersión 0.5.



Función Real vs Función con Ruido



Funciones obtenidas con 3 neuronas.



Funciones obtenidas con 5 neuronas.